Square Peg in a Round Hole
Gender Beyond the Binary

Een vierkant blok in een rond gat duwen
(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de
Universiteit Utrecht
op gezag van de
rector magnificus, prof. dr. H.R.B.M. Kummeling,
ingevolge het besluit van het College voor Promoties
in het openbaar te verdedigen op

vrijdag 13 september 2024 des middags te 2.15 uur

door
Miriam Ines Wickham

geboren op 18 september 1993
te Oberhausen, Duitsland

Promotor:
Prof. dr. N. Ellemers

Copromotor:
Dr. F. van Nunspeet

Beoordelingscommissie:
Dr. F.M. Cramwinckel
Prof. dr. C.Y. van Laar
Prof. dr. D.T. Scheepers
Prof. dr. J.M. van der Toorn
Prof. dr. J.B.F. de Wit

Dit proefschrift werd mede mogelijk gemaakt door de Spinozapremie van de Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO), toegekend aan Naomi Ellemers (toekenningsnummer 01.80.104.00).

Table of Contents

Chapter 1
General Introduction: Square Peg in a Round Hole _____ 5

Part 1: Gender-Inclusive Initiatives

Chapter 2
Exploring the Threat Underlying Online Communications about Gender-Inclusive Initiatives _____ 25
Chapter 3
The Role of Gender-Inclusive Initiatives in Changing Beliefs about Gender and Stereotypes about Jobs _____ 51

Part 2: Gender (Self-)Categorisation

Chapter 4
Gender Identification Beyond the Binary and its Consequences for Social Well-Being _____ 111
Chapter 5
Exploring Implicit Associations of Self- and Other-Concepts with Gender _____ 169
Chapter 6
Gender Categorisation Based on Biological and Social Information: Beyond the Gender Binary _____ 205

End of Part 2

Chapter 7
General Discussion: Attempting to Square the Circle? _____ 255
Supplementary Materials _____ 269
Nederlandse Samenvatting _____ 271
References _____ 280
Acknowledgements _____ 305
Curriculum Vitae _____ 308
KLI Dissertation Series _____ 309

Chapter 1:
General Introduction
Square Peg in a Round Hole

"You've got your mother in a whirl cause she's not sure if you're a boy or a girl"

Rebel Rebel – David Bowie

Author contributions: Miriam I. Wickham (Writing – original draft; Writing – review and editing); Félice van Nunspeet (Writing – review and editing); Naomi Ellemers (Writing – review and editing)

When I signed my guinea pigs up at a local veterinary clinic, besides having to indicate their biological sex, I was also asked to indicate my own gender. The choices I was given were between male and female. I wondered why they needed to know this, especially since it was not my health that would be assessed, but that of my pets. Chances are, you can think of countless examples in which you have had to fill in your gender on a form, even when it seemed irrelevant to do so. Furthermore, chances are that in most of those examples, you were not given more than two choices for your gender. The status quo is a society in which gender is seen as binary: there are only "men" and "women".

In recent years, however, there has been a slow and steady shift in this status quo, whereby we are increasingly acknowledging that there are more than two ways for people to define their gender. Shortly before I began my PhD, for instance, the national rail service of the Netherlands (the "Nederlandse Spoorwegen") announced that they would be addressing their customers as 'passengers' ("reizigers" in Dutch), rather than as 'ladies and gentlemen'. On a global scale, various countries have started to legally recognise "third" gender and sex categories (Ryan, 2018). Gender-neutral pronouns are being introduced to languages that did not yet have them (Gustafsson Sendén et al., 2015) and pre-existing gender-neutral pronouns have become more visible (Bradley et al., 2019). Advocates are arguing for more gender-neutral toilets (Bovens & Marcoci, 2023) and there has been an increased interest in "genderless" or unisex fashion (Kim et

al., 2022). These developments signal a growing awareness of the limitations of the gender binary, which I observed when I was thinking about the topic of my PhD.

The *gender binary* is the view that humans all fit into one of two categories, men and women, that these categories are biologically meaningful, and that this view is natural and unmalleable (Hyde et al., 2019). The effects of this view are pervasive to anyone who identifies outside of the binary, but also for all other LGBT+ individuals (Garelick et al., 2017; Prusaczyk & Hodson, 2020). Recent estimates suggest that anywhere from 3% to 35% of the general population see themselves as part man and part woman, or as something completely different from man and woman (in the Netherlands: Kuyper & Wijsen, 2014; in Flanders, Belgium: Van Caenegem et al., 2015; in Israel: Joel et al., 2014). This suggests that the gender binary may not adequately represent the self-views of a considerable number of people and that a move away from it may be justified. In other words, if many people feel like proverbial "square pegs" being shoved into "round holes", perhaps moving away from those holes is a good idea.

In this thesis, I set out to further investigate the societal move away from the gender binary, and specifically to further challenge two assumptions: (1) that the gender binary is helpful, either as a cognitive heuristic or as a way for us to identify with one another, and (2) that it is natural, in that it is upheld by something innate to humans. To do so, I looked at experiences of threat towards gender-inclusive

interventions, people's gender binary beliefs, and how people categorise both themselves and others by gender. In doing so, I have found further information about how we, as a society, can approach the societal shift away from the gender binary, by raising awareness about this data among the general population and reducing resistance to gender-inclusive interventions.

Importantly, though this thesis is written from the perspective of a social psychologist, I draw from literature in gender studies more generally, which includes psychology as well as (among others) sociology, philosophy, and media studies. As Morgenroth and Ryan (2018) have also pointed out, though there is a large body of social psychological research that fits with the theories and findings of gender scholars in other disciplines (e.g., the philosopher Judith Butler), we often do not integrate them. Though this thesis is not multi-disciplinary, I have summarised key literature from other disciplines throughout, and to let this further inform my choices about what to investigate using social psychological methods. Below, I use this key literature, as well as findings from my own thesis, to challenge the two assumptions that are commonly held about the gender binary: that it is helpful and that it is natural.

Assumption 1: The gender binary system is helpful

Social identity theory (Tajfel, 1974) tells us that social categories emerge from people wanting to identify their in-group to know whom they belong to, whom to favour in times of scarcity, and to provide

them with (information about) group norms (see also Turner, 1987). In that sense, ingroups help us navigate the world by providing us with guidelines for how to behave and improving our well-being by giving us a sense of belonging. While the theory acknowledges that people can belong to multiple different groups and have complex identities, in any specific situation social groups only perform these functions if it is clear who is a member of a certain group and who is not (i.e., groups need clear boundaries). In reality, social group boundaries are not always defined by objective or invariant criteria. For instance, mixed-race people are members of two or more racial groups, and age is a continuum which can be categorised into different age groups.

It seems most people are able to navigate the complexity of social categories. However, the gender binary system perpetuates the idea that gender comes in two "hard-set" categories: men and women. This may be a useful cognitive heuristic for rapid categorisation of in-group and out-group members but given that gender identities are more complex than that, it is also curious that the gender binary is enforced so strictly in society. I thus wanted to further explore this enforcement. For instance, in Chapter 2 of this thesis, I explored one channel through which the gender binary is currently enforced: in public discourse online. The shift away from the gender binary is a sensitive topic that many people are discussing online. Since it represents a large-scale social change, it may also feel threatening to many people. In Chapter 2, I found preliminary evidence that online discourse about

gender, which seems to be overwhelmingly pro-binary, may actually reflect the views expressed by a minority of people - who feel threatened by the notion of gender being non-binary - and not that of the majority. Thus, a minority may perceive the gender binary to be helpful, while others could just as well do without.

In fact, the gender binary may actually be harmful. For instance, in this thesis (Chapter 4), I explored whether gender identification, defined as self-categorisation as a member of a gender group (Turner & Reynolds, 1987) is mostly binary, meaning people identify with one gender group but not with another. I found that a large proportion of the population identified with multiple genders. This goes against the idea of gender being mostly binary. Importantly, I found that people who identified with multiple genders often felt less included by society and therefore their social well-being was suffering from the strict gender binary system.

Furthermore, the gender binary may not only be harmful for those who identify outside the binary, but rather, everyone. That is, in another set of studies (Chapter 3), I looked at how gender-inclusive messaging, as opposed to gender binary messaging, may influence people's beliefs about gender, gender essentialism, and stereotypes about gendered professions. I found preliminary evidence that the gender binary may be reinforcing gender stereotypes and biases, as compared to gender-inclusive messaging.

Lastly, moving away from the gender binary may actually help us navigate the complex gender world better. In another set of studies (Chapter 6), I explored how people deal with, for instance, categorising men who wear women's clothing or people who look physically androgynous (i.e., gender non-conforming people) into binary genders, as opposed to having more than two gender categories to choose from. I found that participants in the binary condition spent more time and energy making a choice. On the other hand, participants in the condition with more than two gender choices often made use of the non-binary option; it offered them a way to categorise someone who would have otherwise been difficult to categorise. As such, having more than two gender choices may actually be quite helpful to people in a world where an increasing number of people are expressing their gender in non-conforming ways.

Assumption 2: The gender binary system is natural

While this thesis is about gender, it is important to also acknowledge the question of "sex", which is argued to be distinct from gender in that it is biological, while gender is social (see Hyde et al., 2019, for a critique that is beyond the scope of this thesis). Sex is argued to be mostly binary (though an estimated 2% of the population are intersex), so a general assumption is that we got our binary gender system from binary sex. Garfinkel (1967) theorized that beliefs about gender being binary are heavily tied to beliefs about gender categories being natural and invariant. Evidence from other primates show that they mostly

separate themselves by sex, fulfilling different roles and even having preferences for human toys fitting with their sex (Alexander & Hines, 2002), showing that there is something innate to gender roles.

However, there are reasons to believe that the current binary gender system is not simply derived from the two biological sexes. For instance, ethnographic studies (e.g., Kessler & McKenna, 1985) have found that features indicating biological sex (e.g., genitals) plays a more minor part in gender categorisation than previously thought. Butler (2002) theorised that our binary gender system does not stem from our belief about sex, but is actually a power play to keep men and women in their social places. She argued that biological sex is used as a tool to justify the gender system.

There is also evidence for the historical and cultural influences on our binary gender system. Historically, the binary gender system was not always in place as it is now. Asian populations often believed in three or more genders, many Native American tribes had up to five gender categories, and even in Europe (e.g., England, Italy, Albania) there were more categories than two (Vincent & Manzano, 2017). The division of labour during the industrial revolution, Christianity and colonisation stemming from Europe to the rest of the world were all leading factors for our (almost) universal binary gender system today (Estrada, 2011; Zlotnick, 2001).

Rather than solely being upheld by innate beliefs about sex and gender, the gender binary is a self-perpetuating belief (Saguy et

al., 2021). The gender binary is upheld and simultaneously upholds the idea that one gender group is superior to the other in a certain domain (gender biases) and subsequent exclusion of one gender group from a certain domain (gender discrimination). Rather than a *natural* tendency to want to place others into one of two rigid categories, researchers have argued that the gender binary is culturally *learnt*, as evidence from developmental psychology suggests (summarised by Hyde et al., 2019). The gender binary is learnt from a young age, for instance, when teachers implicitly treat boys and girls differently (e.g., Duffy et al., 2001; Tsouroufli, 2002), and binary gender labels on toys affect children's play behavior (Yeung & Wong, 2018). As adults, when asked to indicate gender in registration of person characteristics, most forms provide only two options (Morgenroth & Ryan, 2018) and the binary gender categories are considered biologically essential and meaningful (Skewes et al., 2018).

Beauvoir (2000) famously wrote that "one is not born a woman but rather becomes one". With this she meant that assuming one's gender role is not something that happens due to genetics but is something you learn from your social environment. When Bem formulated her sex role inventory (1974), she lived in a world where it was assumed that gendered personality traits were binary and innate. Thanks to her work, a large body of research has investigated psychological androgyny in terms of personality traits. This has impacted gender research greatly, by promoting equality between men and women, showing the malleability of gender roles, and underlining

the negative effects of polarization of gender groups (Dean & Tate, 2017). Still now, women are typically stereotyped to be caring and tender, while men are typically thought of as agentic and dominant, and these stereotypes have persisted across decades (Haines & Stroessner, 2019).

In this thesis, I found further evidence that the gender binary is, at least to some extent, a cultural rather than a natural phenomenon. For instance, in one study (Chapter 2), I found that the online public discourse about gender may be led by a minority of people who are resistant to the shift in gender conceptualization. Our preliminary findings suggest that even those who do not belong to this minority may not be struggling with this shift. In fact, throughout this thesis, I repeatedly make the conclusion that most people find this shift not that problematic or even helpful in various ways. In a world where binary gender is naturally encoded into most of us, you might expect to see universal resistance to changes in the gender binary system, but that is not what we observe in our data.

When I researched people's self-perceptions (Chapters 3 and 4), I found that large proportions of people identified in non-conforming ways or saw themselves in ways that were not strictly binary (i.e., *either* fully man *or* fully woman). These findings suggest that people's perceptions of their own genders are not as rigid as the gender binary system would have us believe. In a naturally binary world, on the other hand, you would expect to find that people's self-perceptions were much more rigidly binary.

In another series of studies (Chapter 6), I investigated gender categorisation of others based on both biological (faces) as well as social cues (clothes) and found that participants' decision-making was sometimes based more on the social than the biological information. This shows that participants do not always rely solely on biological information but that they consider gender to also be socially based. This was evident both from their behaviour (the choices they made) as well as their attention to specific pieces of information (measured using eye-tracking). In the same set of studies, I found that participants made use of non-binary choices when these were provided, depending on context, which showed that participants are aware of the societal tendency to move away from strict binary categories and were willing to engage with non-binary choices.

Social Psychological Research

The works of scholars in other disciplines within gender studies continually inspired me within this PhD, particularly since much of the research conducted within social psychology is done from a binary perspective (Morgenroth & Ryan, 2018). In that sense, our field has much to learn from other fields, and I hope to see more of my fellow social psychologists drawing from other disciplines to inspire them. That being said, it is also my firm belief that social psychology can provide an important contribution to the gender studies landscape, particularly because of the unique approach and methodologies we tend to use in our research. I will outline some of these and mention

how I implemented them in the research in this thesis.

First, social psychologists can take both a top-down and a bottom-up perspective: we investigate how larger social structures shape individuals' experiences, as well as how individuals' ideas and behaviours may affect the larger social structure. Top-down effects can, for instance, be investigated by creating an experimental set-up with different conditions (representing different social structures) and seeing how this affects people's ideas or behaviours. For instance, in Chapters 2 and 5 I developed conditions in which either a binary gender structure or a non-binary gender structure was introduced to participants. Bottom-up effects, on the other hand can be examined by measuring how people feel about a certain topic (e.g., gender identification) in order to understand the larger existing social structure (see Chapters 1 and 3).

Furthermore, social psychologists investigate concepts both explicitly as well as implicitly and can thus gain insight into attitudes that people are aware of, as well as their unconscious biases or involuntary beliefs. Pervasive institutions such as the gender binary can affect both people's surface- and deep-level ideas and it is thus of added value to incorporate both explicit and implicit methodological approaches. This can be done quite literally, by using explicit and implicit measures in a study (e.g., a self-report measure as well as a reaction time task, see Chapters 2 and 4), or somewhat more indirectly, by carefully designing a study in order to frame explicit

measure as a way to capture something that may be more implicit (see Chapter 2).

Moreover, in social psychology we study social constructs from both emotional as well as cognitive perspectives: we understand that a belief may be related to how we feel about a topic or the way in which we process information. For instance, a person's ideas about gender-neutral language may be related to how they feel about their own identity (emotion), or how they process gender-neutral words (cognition). Each of these psychological pathways invites different approaches to solving the issue and increasing support for gender-neutral language. In the chapters in this thesis, we considered both emotional as well as cognitive aspects that may be relevant.

I also looked at people's gendered beliefs about both the 'self' and the 'other', or individual identity and how we see the social group: a distinction that is embedded in social psychological theory. According to social psychological theory, while our self-perceptions are affected by our group memberships and vice versa, the two can also be distinct, and this can change depending on the social context. For instance, in times of stability, an individual may see their group as relatively heterogeneous and themselves as unique members of their group. However, when their group is threatened, they may stereotype other members of their group to make it seem more homogeneous. On the other hand, when an individual identity is threatened, the individual may engage in self-stereotyping to prove that they are a prototypical

member of their group. In this thesis, I kept this distinction in mind and investigated both how people perceive and feel about their own gender, as well as how they perceive and feel about other people's genders (and in Chapter 5, I directly compared the two).

Thesis Overview

The research in the present dissertation was conducted in order to increase our understanding of the societal shift away from the gender binary and towards a more inclusive conceptualisation of gender. In this section, I provide an outline of the chapters within.

While all chapters contribute to investigating the two assumptions (that the gender binary is helpful, and that it is natural), the chapters in Part 1 (Chapters 1 and 2) are somewhat more focused on questioning whether moving away from the gender binary may be more helpful than keeping the binary status quo. The chapters in Part 2 (Chapters 3, 4, and 5), on the other hand, are more focused on questioning whether the gender binary is perhaps more socially constructed than one might assume based on the dominant narrative.

Part 1: Gender-inclusive initiatives

The thesis begins with gender-inclusive initiatives as a starting point, which are initiatives designed to be inclusive of people on the whole gender spectrum. As mentioned at the beginning, when I started my PhD, I observed many gender-inclusive initiatives being implemented on a broad scale. I observed how the media discussed

these initiatives in articles, as well as the discussions among laypeople reading these articles. I wanted to investigate people's thoughts and feelings underlying these discussions, which I did in Chapter 2. Furthermore, I wanted to investigate what hidden benefits gender-inclusive initiatives might have, besides being more inclusive of non-binary and transgender people, which I did in Chapter 3.

In Chapter 2, I investigated communication about gender-inclusive initiatives from both the perspective of those who engage in online discourse about this topic (the 'writers'), as well as those who may read the online discourse (the 'readers'). Specifically, using qualitative analysis, I explored whether writers' online comments could be located into different kinds of communications of social psychological threat. Furthermore, using a quantitative approach, I investigated how much readers agreed with these comments and what underlying factors may contribute to agreement. The findings provide insight into how beliefs about gender may be perpetuated, by whom, and why.

In Chapter 3, I examined whether and how gender-inclusive initiatives may benefit cisgender men and women by reducing gender stereotypes, particularly regarding job suitability. Across three studies, I investigated how gender-inclusive initiatives influenced gender beliefs, stereotypes about gendered jobs, and interest in these jobs. I did so by presenting participants with either a non-binary (based on real gender-inclusive initiatives) or a binary manipulation and included both explicit

(self-report) and implicit (reaction times) dependent variables. The findings provide insight into hidden positive effects of gender-inclusive initiatives for cisgender individuals.

Part 2: Gender (self-)categorisation

In Part 1, I showed evidence that the gender binary being pervasive was not related to it being helpful, since moving away from the gender binary through gender-inclusive initiatives may actually have many benefits. In Part 2 of the thesis, I then wanted to explore whether the pervasiveness of the gender binary may be related to it being naturally encoded into us in ways that previous literature has given limited consideration to: people's need to identify with and categorise others within groups with clear boundaries. Across three chapters, I thus explored whether people's categorisations of themselves (Chapters 3 and 4) and others (Chapters 4 and 5) were binary or not, and what factors non-binary categorisations may be related to. The findings in these chapters contribute to existing literature about gender non-conformity and gender social cognition.

In Chapter 4, I examined how people categorise themselves in terms of gender and whether people's self-categorisations fit more into a binary model or a non-binary model of gender. I also investigated how gender self-categorisation may relate to general well-being. Across three studies, I asked participants to indicate how much they identified with different (binary and non-binary) genders and about

their well-being on various measures: societal inclusion, self-esteem, life satisfaction and general affect. The findings indicate that people's self-categorisations are not strictly binary, as one might expect in a world where the binary is naturally encoded into us. Furthermore, our findings provide insight into the negative effects of forcing people to conform into a binary self-categorisation on well-being.

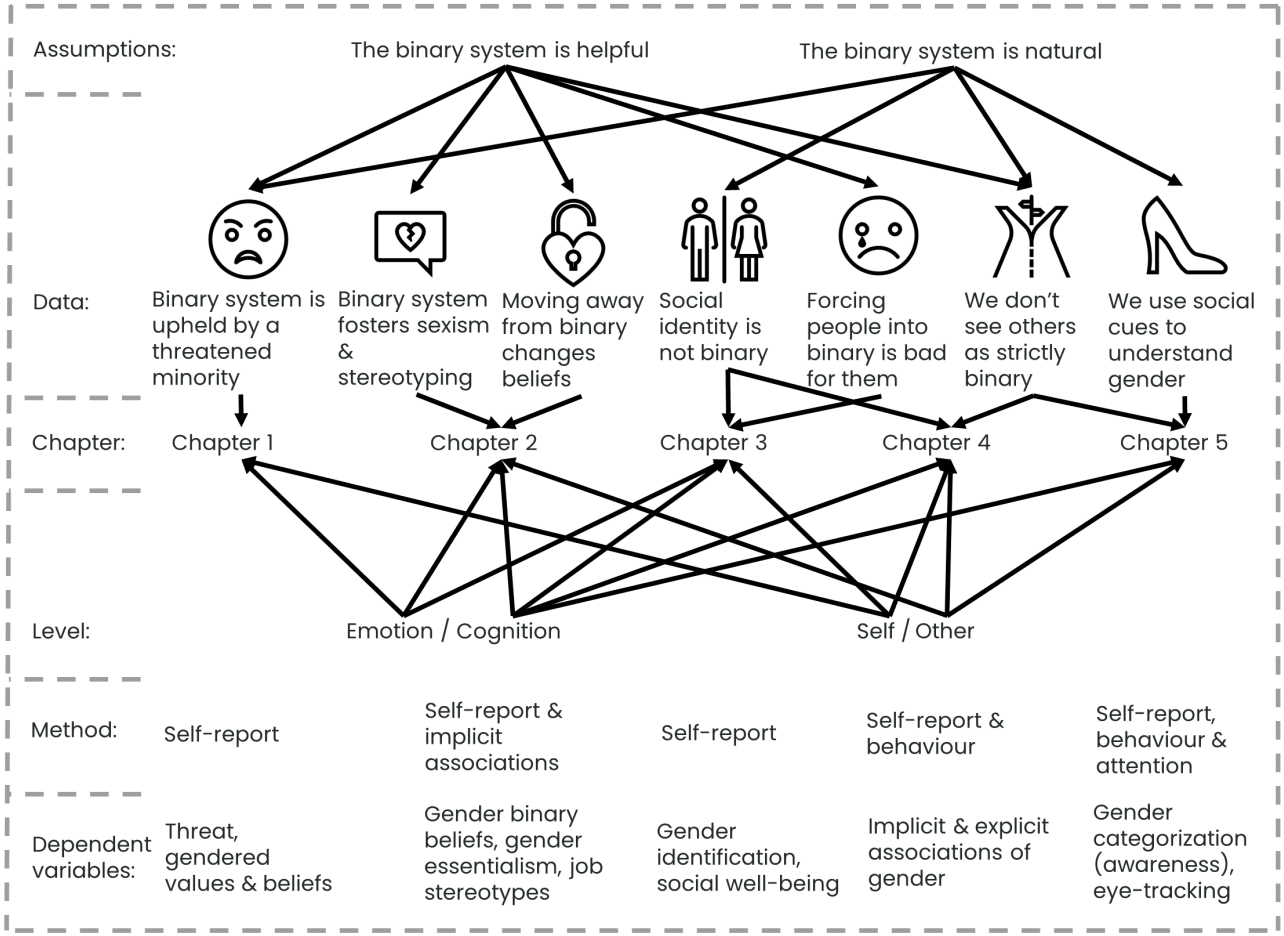
In Chapter 5, I extended the findings of Chapter 4, by measuring self-categorisation implicitly, and additionally exploring implicit categorisations of others (the gender ingroup and outgroup). Using a reaction time task, I showed that implicit self-categorisations follow a similar patterns as explicit ones, thus showing further evidence that people may not naturally identify in a strictly binary way. Furthermore, findings about implicit ingroup and outgroup categorisations provided insight into how these relate to self-categorisations. Our results indicated that people may not see others as strictly binary either, which laid the foundation for my research focus in Chapter 6.

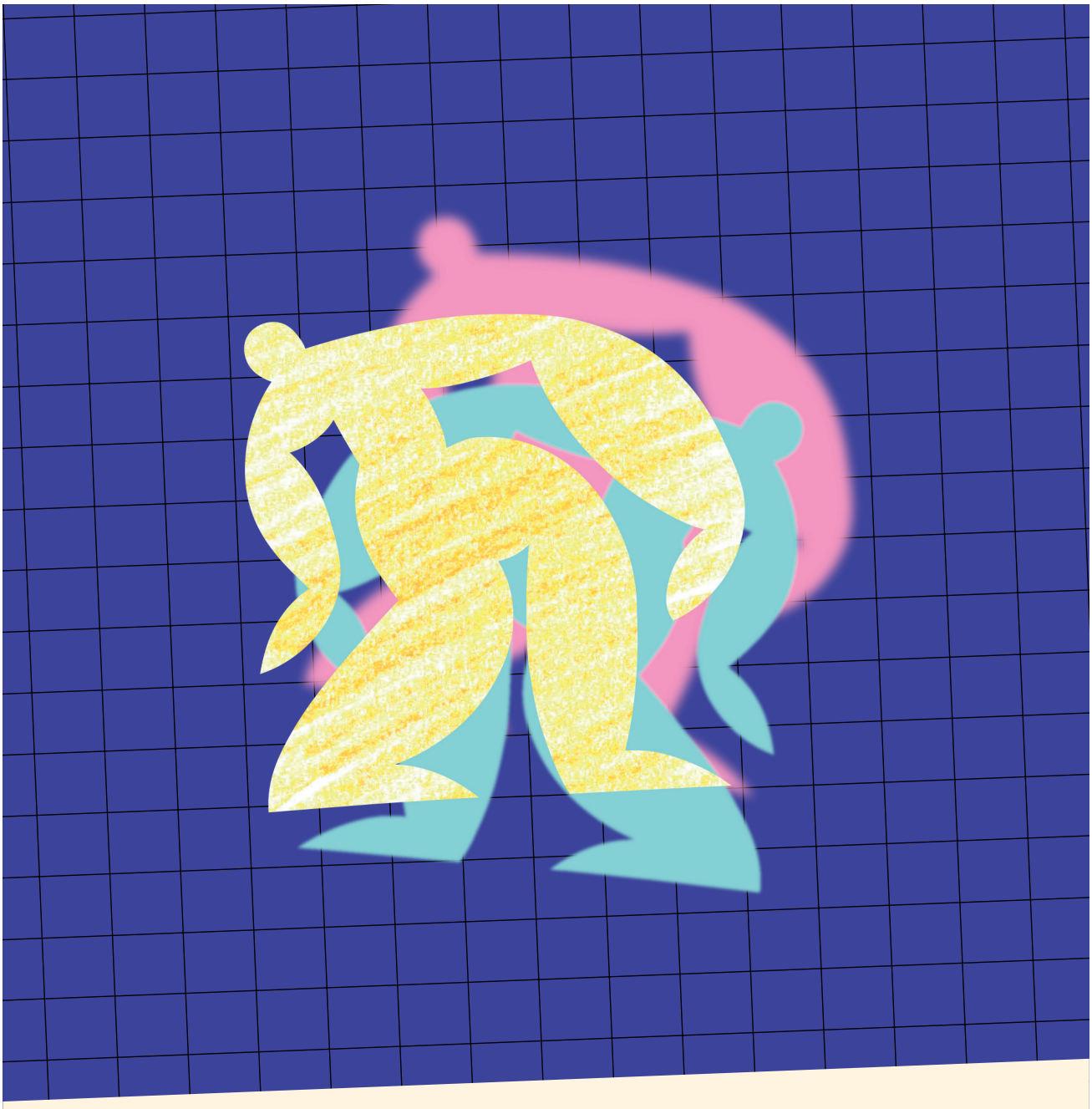
In Chapter 6, I focused on how participants categorised others by gender and whether this more closely followed the narrative that "gender is biological" or the narrative that "gender is socially constructed". Specifically, across four studies, I investigated whether people's gender categorisations were more based on biological cues (e.g., your natural facial structure) or social cues (e.g., the clothes you wear). I also investigated how this may differ across social contexts, where different information (biological or social) may be more relevant

for understanding someone's gender. I investigated this by using both explicit measures (self-report) and implicit measures: behaviour in a categorisation task and attention to biological or social information, as measured through eye-tracking. Our findings provide insight into how people process gender information about others and how this may change as gender-inclusive initiatives become more common.

It should be noted that the chapters in this thesis have been written such that they can be read independently of the other chapters, so the reader may see some overlap or repetition between them.

A visual representation of the findings of each chapter, which assumption they focus on, at which level (emotion/cognition, self/other) mechanisms are addressed, and which methodology was used, can be found on the next page.


Assumptions:
The binary system is helpful
The binary system is natural
Data:
Binary system is upheld by a threatened minority
Binary system fosters sexism & stereotyping
Moving away from binary changes beliefs
Social identity is not binary
Forcing people into binary is bad for them
We don't see others as strictly binary
We use social cues to understand gender
Chapter:
Chapter 1
Chapter 2
Chapter 3
Chapter 4
Chapter 5
Level:
Emotion / Cognition
Self / Other
Method:
Self-report
Self-report & implicit associations
Self-report
Self-report & behaviour
Self-report, behaviour & attention
Dependent variables:
Threat, gendered values & beliefs
Gender binary beliefs, gender essentialism, job stereotypes
Gender identification, social well-being
Implicit & explicit associations of gender
Gender categorization (awareness), eye-tracking



Part 1:
Gender-inclusive initiatives

Chapter 2:
Exploring the Threat Underlying Online Communications about Gender-Inclusive Initiatives

*"The images that f***ed ya*
were a patriarchal structure,
and you never will surrender
to that narrow view of gender"

Black Tie – Grace Petrie

Author contributions: Miriam I. Wickham (Conceptualisation; Data curation; Formal analysis; Investigation; Methodology; Validation; Writing – original draft; Writing – review & editing); Félice van Nunspeet (Conceptualisation; Methodology; Validation; Writing – review & editing); Naomi Ellemers (Conceptualisation; Funding acquisition; Methodology; Validation; Writing – review & editing)

Abstract

Gender-inclusive initiatives, such as gender-neutral bathrooms, spark polarizing online discussions, which can affect public opinion and influence law-makers. In this paper, we investigated online discourse about gender-inclusive initiatives in two phases. In phase 1, using social psychological theory, we explored whether people who post online comments about GII are not only explicitly communicating their opinions, but also implicitly communicating their underlying feelings of threat. Based on social psychological theory, we identified five types of social psychological threat underlying online comments. Next, we developed a measure with which we could explore whether a larger group of people agreed with these comments. In phase 2, we used this measure to explore differences in agreement depending on the type of threat communicated in a comment. First, our results suggest that those who write these comments do not represent the opinions of a larger group of people. Furthermore, our findings suggest that particularly women's threat regarding gender-inclusive initiatives is low, contrary to the popular narrative regarding women's concerns for their safety. Our findings help us to be critical of online arguments against anti-trans* inclusion by understanding underlying motivations and shed light on how resistance to gender-inclusive initiatives may be reduced in the future.

Discussions about *gender-inclusive initiatives* (GII), which are designed to be inclusive of trans*[1] people, have been on the rise throughout the Westernized world (e.g., gender-neutral bathrooms, Blumell et al., 2019). Online discussions about GII among laypeople, for instance in online comments sections, are often polarizing (e.g., Colliver et al., 2019). Furthermore, such comments can affect opinion formation of those who read them (e.g., Lee et al., 2022). In this paper, we explored online discussions about GII in two phases. In phase 1, we made an inventory of the types of comments people write under online newspaper articles about GII. We posit that people who write online comments may not only be communicating their opinions but may also implicitly be communicating their underlying feelings of threat towards GII. It should be noted that, while they may not *explicitly* be communicating that they feel threatened, we still refer to their underlying feelings as 'communications of threat' throughout this paper, for ease of reading. Based on social psychological theory, which states that different types of threat contribute to resistance to social change (e.g., Morgenroth & Ryan, 2021), we found that we could locate comments into five dimensions, based on five types of threat. Next, we developed a measure of threat towards GII based on these comments. In phase 2, we used our measure to investigate how much people agreed with these comments, indicating their own feelings of threat. We also explored how participant gender, and other gender-related beliefs, were related to agreement to such comments. In doing so, we gained insight into the feelings and processes underlying resistance to changes in the gender system. Understanding these underlying feelings is important, so that we can be critical of arguments against trans* inclusion and to learn how to reduce resistance to GII in the future.

Online Discussions about Gender Inclusiveness

In recent years, a heated public debate about gender has gained considerable traction in media coverage. In this debate, one side is

arguing that gender is binary (there are only two genders: men and women) and based on biological sex, while the other side is arguing that gender is a free identity located on a spectrum (Cooper, 2019). We can see this debate in the increase of gender-inclusive initiatives (GII) that are aimed at improving the lives of trans* people, while there is also considerable backlash against these kinds of initiatives (Blumell et al., 2019). Given how polarizing the topic of trans* inclusion is, it is no surprise that many news articles have covered this topic (e.g., McGann, 2017), and thus that there are many online discussions about it.

Notably, online discussions can give a skewed picture of public opinion, since it is often a small group of people with particularly radical views who lead these discussions, much more so than in offline discussions. For instance, those who see themselves as having a minority opinion are more likely to be outspoken about it in online environments, while this would not be the case offline (Porten-Cheé & Eilders, 2015). This is because people feel less inhibitions about what they communicate online, as compared to face-to-face, because of the perceived anonymity and invisibility (Lapidot-Lefler & Barak, 2012). Indeed, it has been found that those with a view that is contrary to common public opinion (as measured by population-wide surveys) tend to represent a vast majority of the contributions of online comments (De Kraker et al., 2014). This also seems to be the case for the topic of trans* inclusion.

Importantly, online discussions about polarizing topics, like gender inclusiveness, can have a substantial influence on public opinion formation. For instance, online comments left under news articles can influence people's views of the credibility of the article (Pjesivac et al., 2018), even though these comments may only reflect the opinions of a small group. Furthermore, online comments sections influence people's views of the public opinion, which they tend to want

to conform to (Lee et al., 2022). Moreover, online comments about polarizing and political topics tend to be very emotional, which further increases their influence (Webster & Albertson, 2022). It is thus relevant to study online discussions about GII, and people's opinions about these online discussions.

Opinions Opposing Trans* Inclusion

One of the most prominent recent examples of media coverage about the public gender debate has been the controversy around J.K. Rowling's views on gender (Duggan, 2022). Rowling has been argued to represent trans-exclusionary radical feminists (TERFs), a group of people who have been very outspoken about their beliefs regarding trans* inclusion (e.g., Yang, 2021). Their anti-trans* argumentation is often centred around the idea that gender cannot be self-defined (Raymond, 1980; Stock, 2021), and that trans* inclusion has negative effects for women's safety (Turnbull-Dugarte & McMillan, 2023) and rights (Jeffreys, 2014; Pearce et al., 2020).

However, those who are most outspoken about being anti-trans* are likely a small group. For instance, it has been found that most of the British public agree that trans* people should have rights such as easier gender recognition and access to gender-specific spaces (Smith, 2020), contrary to anti-trans* ideology. Nonetheless, this small anti-trans* group seems to have been influential in the decision-making process of law-makers (McKinnon, 2018). For instance, anti-trans* arguments played a central role in the so-called US 'bathroom bills', which deny trans* people the right to access toilets that are in line with their gender identity (e.g., Ura, 2015).

When we were contemplating the kinds of arguments made against trans* inclusion, and therefore initiatives that support that (GII), a key word that stuck out to us was 'threat': for instance, the

threat to women's safety and rights. This is because, in our field of study, namely social psychology, the widely recognised theory is that resistance to social change is often driven by (different types of) threat (e.g., Scheepers et al., 2009). For instance, social psychologists argue that people are motivated to maintain the status quo, since this provides certainty and security. They thus feel threatened by change (for instance, to the gender system), even when it would mean positive outcomes (Jost et al., 2010). Thus, it seemed to us that people who are active in online discussions about this topic may be communicating their feelings of threat. We thought we may be able to identify different types of social psychological threat in their comments, which would provide insight into what types of threat to address to reduce resistance to GII in the future. Furthermore, we thought that we could use social psychological research methods to also explore the readers' perspectives: how readers of these comments may feel about them, and how that is related to other beliefs they may hold. In this project, we wanted to gain a better understanding of the feelings and processes underlying resistance to changes in the gender system, which may have practical implications for designing more effective GII in the future.

The present paper

In the present paper, we explored online discussions about trans* inclusion and related these back to social psychological theory and methodology. We did this in two phases. In phase 1, we focused on the people who are active in these online discussions and made an inventory of the types of comments they tend to leave under online news articles about this topic. Using social psychological theory, we qualitatively investigated comments to explore whether and what types of threat they may be communicating. We then developed a measure based on comments that communicate threat, to be used in phase 2.

In phase 2, we focused on the readers' perspective, i.e., people

who may stumble across an online news article about gender inclusion and may see comments underneath it that communicate threat. We used the measure we developed in phase 1, as well as other social psychological instruments, in a survey to explore how much people agree with these communications of threat, and what underlying beliefs their agreement might be related to. Given that people who are outspoken about their opinion online are usually not representative of the public opinion, we expected agreement with these comments to be relatively low. However, exploring agreement with these comments provided us a way to capture implicit feelings of threat that readers may feel towards GII. Furthermore, we explored whether agreement differed depending on gender of the participants, since the narrative seems to be that GII are particularly threatening to women.

Phase 1: Identifying Threat in Online Comments and Developing a Measure

As Morgenroth and Ryan (2021) argue, different types of social psychological threat may contribute to resistance to trans* inclusion. The inclusion of trans* people, by means of gender-inclusive initiatives or laws, represents a disruption in the current social system, where people are categorised almost exclusively as men or women based on the sex they were assigned at birth. According to social psychological theory, disruptions to the social system are often related to threat, because they represent a shift in the status quo, a shift in the social categories and identities we are attached to, or a shift in power dynamics.

In this phase, we were inspired by Vergoossen et al. (2020) who researched criticism against gender-fair language (one type of GII). When reading their paper, we thought that many of the quotes from their participants felt like communications of social threat towards losing the gender binary (Morgenroth & Ryan, 2021). For instance, there were arguments defending the linguistic status quo, where participants appealed to tradition, argued that change is too difficult and justified the

current system with appeals to authority. There were arguments about gender-neutral language removing the distinction between men and women, which could be seen as a threat of losing the social categories and identities they are attached to. The authors also described worries about women's rights and justifications of misogynistic language, which may be communicating threat of losing existing power dynamics (where men have more status than women and women have fought hard for the status they have). When we read their paper, we realised we could identify social psychological threat in many of the quotes, and we thus thought we may be able to do the same in a qualitative analysis of online comments sections. Doing so would provide insight into what types of threat underly negative opinions about GII and how we may be able to reduce resistance to GII in the future.

We thus made inventory of comments and locate them into different types of threat. We searched for recent articles on the Dutch online newspaper nu.nl. We chose this online newspaper because it is considered to be neutral (not related to a particular political party), it reaches a large percentage of the adolescent and adult population of the Netherlands (NOM bereiksonderzoek, n.d.), and it is an online newspaper where comments under an article are common. We used the search terms "gender", "neutraal" ("neutral"), and "non-binair" ("non-binary") and picked articles that mentioned some form of GII that were relatively recent (January 2020 onward). Our analysis of various online comments was deductive because we used social psychological theory to understand the types of threat commenters may be communicating. It was thematic because we located comments into themes based on theory, which were different types of social psychological threat. We took a pragmatic approach and adopted a critical realist ontology and epistemology. This is because we believe that the knowledge gained from the online comments reflects a psychological reality of these

online commenters. We describe each of the types of communications of threat that we found, in turn, below. We start by describing system threat because it is an overarching type of threat. Next, we describe distinctiveness and safety threat, which are types of intergroup threat, meaning they relate to how gender groups perceive and interact with each other. Last, we describe status threat, for which we differentiate between men's and women's status.

Five Types of Threat in Online Comments

We found examples of comments that seemed to communicate system threat, for instance, by justifying the gender binary system because it fits the majority of people's gender identities. According to theory, existing systems provide us with guidelines for how to interact with one another and therefore reduce uncertainty. As such, some people may be motivated to justify the (gender) system to cope with the status quo and their place in it, and may feel threatened by changes to the system (Jost et al., 2010; Prusaczyk & Hodson, 2020). An example of communication of system threat was an argument justifying the binary system by calling non-binary people "overly sensitive" and saying that a vast majority of people are binary (J.K. Rowling Nuanceert Uitspraken, 2020).[2]

We also found examples of communications of distinctiveness threat in online comments, for instance, by expressing their discomfort with not being addressed as male or female. According to psychological theory (Ellemers et al., 1999), members of social groups have a need for clear distinctions between social groups. Thus, some gender-neutral initiatives may feel threatening to people because the distinction between gender groups is removed. Indeed, distinctiveness threat has been found to lower support for GII (Hayes & Reiman, 2022). An example of a communication of distinctiveness threat was a comment discussing issues with gender-neutral language and toilets,

saying "I still want to be addressed as lady or Mrs." and expressing that her identity is important to her (Marrewijk & Einder, 2021).[3]

There were also communications of safety threat in online comments, whereby commenters argued against GII by citing that men would use these initiatives as a way to endanger women, in line with anti-trans* rhetoric. Indeed, a majority of women have experienced some form of sexual harassment since the age of 15 (Skinner, 2018); it is thus no wonder that women may be scared for their physical safety. However, it should be noted that there is insufficient evidence to argue that trans* people, or even cisgender men pretending to be trans*, are any significant proportion of the perpetrators in cases of sexual harassment against women (Bianco, 2015). Nevertheless, safety threat is sometimes used by women to argue for social change (Nelson et al., 2008) and sometimes against social change (Outten et al., 2019). Importantly, arguments against social change based on women's safety may not only be endorsed by some women, but also by men who believe that women should be protected (Glick & Fiske, 1996; Morgenroth & Ryan, 2021). An example of communications of safety threat was that it would be naïve to think that gender-neutral toilets would be a safe space for women because "there will always be perverts who take the opportunity to harass women" (Aparte Wc's, 2020).[4]

Next, we found communications of men's status threat in online comments, such that commenters expressed their need for traditional gender roles whereby men have higher status than women. According to literature, members of higher status groups (e.g., cisgender men) may feel threatened if they perceive a shift in the power relations between men and women (e.g., Scheepers et al., 2009). GII represent a potential shift in the power relations between men and women since they represent a dismantling of the gender system as a whole (Morgenroth & Ryan, 2021). Men's status threat may also be related to the idea that

manhood is more precarious (or 'fragile') than womanhood, in that it can be lost if men do not continually prove themselves (Bosson et al., 2013). Men's status threat may therefore express itself in appeals to simpler times, 'when men were still men'. An example of communication of men's status threat included "I still remember how simple it used to be" (Marrewijk & Einder, 2021).[5]

Lastly, there were indications of communications of a different kind of status threat, namely women's, such that women expressed their concern for decreasing attention towards women's rights as attention towards trans* rights increased. Women have had to fight for their status and rights, which have been steadily increasing for several decades (Malinowska, 2020). A lower status group, namely trans* people, dismantling the gender system in which women have gained those rights, may thus elicit a type of status threat in women. Indeed, evidence suggests that stronger feminist identity (and therefore worry about women's rights) is related to more stigma towards trans* people (Worthen, 2022). An example of communication of women's status threat included "greetings from a group that I don't hear from anymore because of these attention seekers", implying that women's (and other groups') rights are being undermined by attention towards trans* rights (Marrewijk & Einder, 2021).[6]

We thus concluded that we could indeed locate comments under online articles about GII into different themes of social psychological threat. We believe, based on theory and our findings, that it is likely that people who are active in online discussions about GII, particularly those who voice their strong opinions about them, are feeling and communicating types of social psychological threat. Next, we wanted to develop a measure, based on our findings, in order to investigate agreement with online comments in people who read them, which we did in phase 2.

Development of a Measure of Threat Towards GII

Based on quotes found in comments under real online articles and quotes reported by Vergoossen et al. (2020), we developed a set of 15 comments, such that a set of three would communicate one of five kinds of threat. We then piloted them among 7 people for how credible they were and made appropriate changes to our items based on this pilot. For more information about this pilot, see Supplementary Materials. Our final items, which we visually designed according to the comment section of the British online newspaper Metro (metro.co.uk; see Figure 1) for the purposes of placing them under a fake news article from Metro in phase 2, can be found in Table 1. We thought this measure would reflect the way that many people interact with information these days, namely online content and comments sections where people share their opinions and use this information to form their own opinions.

Figure 1

Example of a Stylized Comment

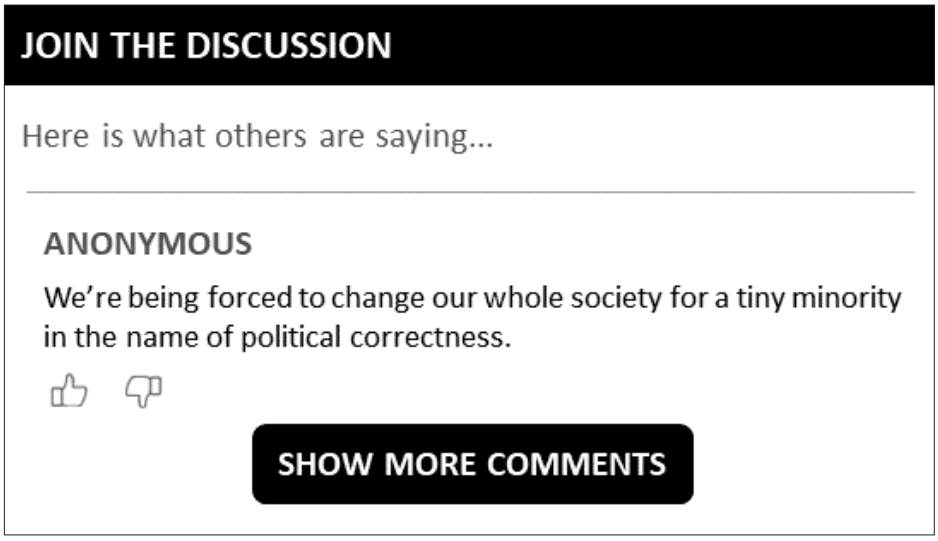


Table 1

Final Comments for our Measure of Agreement with Communications of Threat

Item	Type of threat
"We're being forced to change our whole society for a tiny minority in the name of political correctness."	System threat
"I see no reason to change our whole language for a few people."	System threat
"Why do we need a whole new clothing line for a few 'special snowflakes'? I just don't get it"	System threat
"I want to be seen as a man or a woman, not a genderless thing."	Distinctiveness threat
"Gender-neutral language feels too impersonal, I don't want to be spoken to like that."	Distinctiveness threat
"What about those of us who don't identify as attack helicopters? I think we should still just be called men and women."	Distinctiveness threat
"Perverts will use this opportunity to harass women in dressing rooms by pretending to be female."	Safety threat
"I see this as a danger to all the women in my life. Men are opportunistic – you can't change that."	Safety threat
"Biological men are going to take advantage of women and we're just gonna let it happen so we don't "offend" anyone? This is nonsense!"	Safety threat

“Things used to be simpler: men were men and they provided for their ladies – what’s wrong with that?”	Men’s status threat
“I’m just a regular guy who likes girls, but apparently that makes me the devil nowadays.”	Men’s status threat
“Men can’t just be men anymore – they have to be forced to wear dresses, apparently.”	Men’s status threat
“We need to focus on women’s rights, not this stuff.”	Women’s status threat
“I never hear about women’s issues anymore because of these whining attention seekers.”	Women’s status threat
“All these policies are just distracting from the real issues. Women are still being paid way less than men and I’m supposed to care about this?”	Women’s status threat

Phase 2: Investigating Agreement with Online Comments

In this phase, our aim was to explore to what extent a larger group of people shared the same sentiments as those who write comments under online articles about GII, and whether they did so differently for different types of underlying threat that we uncovered. To explore a larger group of people's implicit feelings of threat, we did not focus on previously validated measures of experienced threat, since such measures often require a lot of reflection from the participant. Agreeing with another person's comment, on the other hand, may be more intuitive and closer to how people encounter and process opinions about GII in daily life. Thus, we measured to what extent people agreed with the different comments which implicitly communicated different types of threat and posit that agreement with these comments may indicate that participants also felt threatened themselves. Given that online comments might be more extreme than typical items in previously existing, validated questionnaires, we thought that agreement with these comments may be rather low. Nevertheless, we think our measure captures whether people share certain sentiments in a more implicit way.

We were also interested in how agreement with these comments may be related to other social psychological measures, such as beliefs and opinions about gender, which have often been found to correlate with social psychological threat (Morgenroth & Ryan, 2021). Furthermore, we wanted to explore whether agreement with these online comments may differ depending on the gender of the participant, since men and women often report differences in threat and other gender beliefs (e.g., Jost & Kay, 2005). Lastly, we wanted to see whether people's agreement with these online comments would differ depending on the article they were placed under. As such, one article described an identity-blind and the other an identity-conscious

GII, since it has been shown that people tend to feel differently about these two types of interventions (Morgenroth et al., 2021).

To explore this, 159 participants were randomly assigned to read one of two news articles about a GII (Appendices A and B).[7] Next, participants rated, on a 7-point Likert scale, how much they agreed with each of the comments that we developed in phase 1.[8] Then, participants answered questions about social psychological constructs that have been found to be related to threat, such as gender system justification: the belief that the gender system in society is fair, which has been related to system threat (e.g., Mallett et al., 2011). Lastly, participants answered questions about their demographics. Our sample consisted of a majority of cisgender women (68.55%), mostly white (94.34%), not disabled (88.05%), and a sizeable minority were LGBTQIA+ (18.87%). Participants' mean age was 22.99 (SD = 5.05). Participants were recruited among Psychology Bachelor students of Utrecht University and on social media.[9]

All our analyses were exploratory, and we summarize some of our key findings, which have potential theoretical and practical implications. First, we found that, generally, our participants did not agree with the online comments, as one might expect. This is further evidence that communications of threat about GII in online comments may not reflect how a larger group of people feel about GII. Second, we found interesting differences between the men and women in our sample. Our findings suggest that women in particular, who represented a majority in our sample and who tended to be quite young, did not agree that GII have negative implications for their safety or their rights, suggesting that anti-trans* argumentation does not reflect a reality of young women's concerns. Instead, our male participants tended to report higher agreement with the comments, suggesting that they felt more threatened by GII. Interestingly, men who were more paternalistic

tended to agree more with comments suggesting that GII threaten women's safety, suggesting that anti-trans* argumentation is in line with men's ideas that women must be protected.

Low Agreement with Online Communications of Threat

Our first key finding was that our participants' agreement with the online comments was overall low (M = 2.44). This may be partly due to our sample being relatively progressive, given that many of them were recruited among highly educated university students. However, it is also in line with previous literature, showing that opinions shared online, for instance, in comments sections, are usually somewhat extremist and do not reflect the opinions of the general public (Porten-Cheé & Eilders, 2015). Our finding is a further indication that implementing GII may not be as threatening as online discussions might have us believe. This also has practical implications for practitioners and lawmakers, who may want to take news articles and online discussions with a grain of salt when making important decisions about trans* inclusion.

Differences in Agreement between Gender Groups

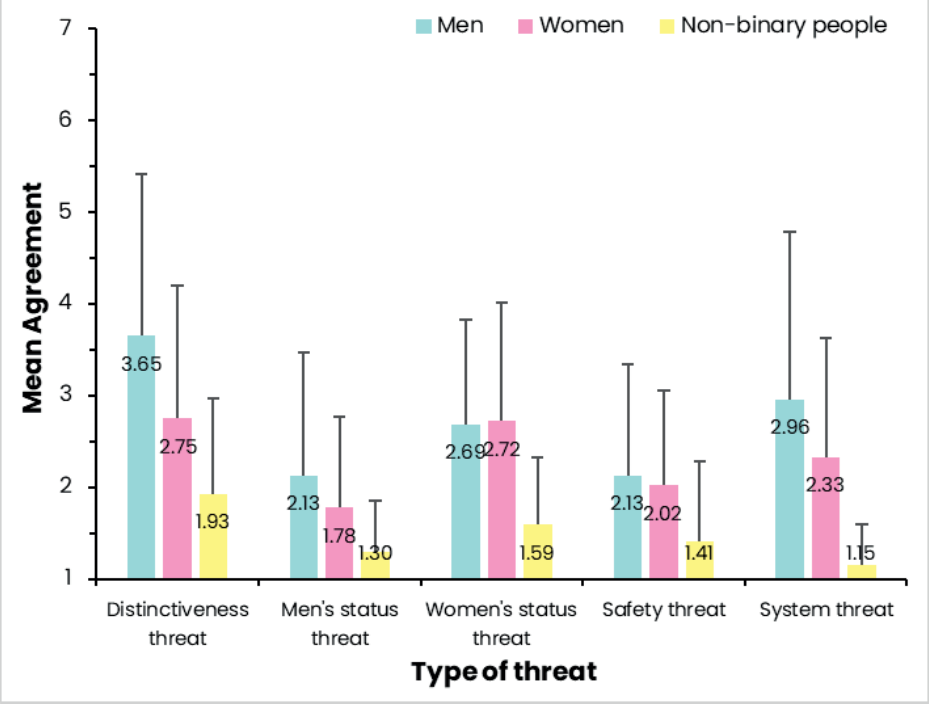
Next, we explored differences in how much people agreed with our online comments, depending on their gender group. We found that our very small sample of non-binary people (nine participants) tended to report very low agreement with the comments, and substantially lower agreement than the women and men in our sample (for descriptive statistics, see Figure 2). This is unsurprising, given that GII are designed to increase inclusion of non-binary people. Due to the very small sample size of non-binary people (N = 9), we did not investigate their agreement further.

Furthermore, women's agreement with arguments regarding women's status or safety threat was rather low. This is contrary to anti-trans* argumentation, which is often focused on concern for

women. Interestingly, men tended to agree more than women with the comments in general, including those concerning women's safety. We thus decided to explore men's agreement further, by looking at how their agreement was related to other beliefs that they held.

Figure 2

Descriptive Statistics of Agreement with Different Types of Comments Per Gender Group



Note. Error bars represent the (positive) standard deviation of the mean.

We found that men's agreement with comments communicating concerns for women's safety was particularly pronounced in men who reported more benevolently sexist beliefs. Benevolent sexism (Glick & Fiske, 1996) is the set of attitudes that see women as justifiably weaker and lower status, but that are subjectively positive (for those who hold these attitudes), in that they see women as people that

must be cherished and protected. Our finding that agreement with concerns for women's safety was related to men's benevolent sexism is in line with literature (Morgenroth & Ryan, 2021) and unsurprising. Benevolent sexism is a form of paternalism of men towards women, and benevolently sexist behaviour by men has been found to have substantial negative effects for women (Dardenne et al., 2007). We may thus need to be critical of resistance towards trans* inclusion that is based on the concern for women's safety, since those concerns may be based on benevolently sexist beliefs.

Our findings also suggested that men agreed more with comments communicating distinctiveness threat, men's status threat, and system threat. We found that these three constructs converged with one another to a large extent. The gender system in society is one where men have higher status, and where the distinction between men and women is made perpetually salient. Thus, our society is a gender binary system (Hunt, 2018) as well as a patriarchal system (Segato & Monque, 2021), and men's higher status and distinctiveness from women are embedded into the system. It thus makes sense that men who agree with concerns about our political and social system being dismantled (system threat) would also agree with concerns of men losing their higher status in society (men's status threat) as well as concerns of the distinction between men and women being lost (distinctiveness threat).

Men's agreement with communications of threat was related to hostile sexism and gender system justification. Hostile sexism is the set of attitudes towards women whereby they are seen as exaggerating their problems and pretending to be discriminated against (Glick & Fiske, 1996). Hostile sexism is highly related to gender system justification, which is the belief that the current patriarchal system is justified, in that men have just as many chances to succeed as women (Jost & Kay, 2005). It is easy to see how these two beliefs are related: one

cannot justify the current gender system without believing that people who complain about the system (e.g., women) are exaggerating their problems. The justification of the system, and thus negative attitudes toward those who are unhappy with it, is related to concerns about the current system being dismantled (i.e., system threat), and thus with concerns about men's status and the distinction between men and women, since these are embedded in the system.

Furthermore, men's agreement with communications of threat was related to gender essentialist beliefs. Gender essentialism is the belief that men and women are fundamentally different, and that these differences are based on immutable biological features (e.g., one's DNA; Skewes et al., 2018). Those who hold gender essentialist beliefs are likely concerned about losing the distinction between men and women (i.e., distinctiveness threat), explaining our findings. Furthermore, people with gender essentialist beliefs tend to use biological differences between men and women to justify differences in status between men and women (thus, men's status threat; Fine, 2017), and thus the patriarchal gender system (system threat). Our findings give us an indication into the beliefs and thoughts that may underlie men's threat towards trans* inclusion and help us to be critical of concerns for distinctiveness, men's higher status and the current system.

Discussion

In this paper, we have reported our mixed methods research, which we conducted in two phases. In phase 1, we focused on the perspective of those who leave negative online comments under articles about gender-inclusive initiatives (GII) and found that their comments could be located into different kinds of communications of social psychological threat. We also developed a measure based on our qualitative investigation into 'real' online comments, for use in phase 2. In phase 2, we used our measure, as well as other

social psychological measures, to provide insight into the reader's perspective, rather than the writer's. Our findings from phase 2 suggest that negative online comments about GII may not reflect common public opinion about GII. Furthermore, we showed how agreement with online communications of threat may differ between men and women, and which underlying factors may contribute to agreement. Specifically, we found that men tended to agree more with communications of threat, and thus provide an alternative perspective to the idea that women are concerned for their safety and rights with regards to GII. These findings help us to be critical of arguments against GII. Our findings also suggest that practitioners and lawmakers should keep in mind that online discussions about GII may reflect the opinions of a small group and not everyone may feel threatened by GII. Throughout the paper, we have discussed the theoretical and practical implications of our findings.

Limitations and Recommendations for Future Research

Our research was a first step towards a better understanding of a) types of threat towards GII, b) how that threat is communicated in online environments, and c) factors that contribute to people's agreement with online communications of threat. However, there were certain limitations to our studies. First, our qualitative investigation of online comments was restricted: we only looked at articles from one online newspaper, which were relatively recent. A more extensive systematic analysis of online comments is recommended for future research, because it would provide a more complete picture. Second, our quantitative investigation of reader's agreement included a limited sample of participants, such that they were largely young, white, and cisgender. Younger people are likely to hold more progressive views (Central Bureau for Statistics, 2019), while a majority white and cisgender sample misses the important perspectives of POC and

trans* people, which likely differ from the majority groups (Kasai, 2023). It is thus recommended that our research is replicated among a more diverse sample.

Lastly, in this research we focused on the negative opinions, beliefs and effects of GII, which helped us understand resistance towards these initiatives. However, there may also be positive effects of GII, such that a dismantling of the current gender system may reduce sexism, transphobia, homophobia, and other types of discrimination. In future research, it is thus important to research how using GII may positively affect people's beliefs regarding gender equality, for both cis-heterosexual individuals as well as members of the LGBTQIA+ community.

About the Authors

The authors have reflected on how their positionality affects their views about this topic, and thus how their subjectivity may have affected the arguments made in this paper. Three of the authors are white, while one is of Jewish descent, which may affect our views of gender since colonialism and religion have played a central role in the formation of the gender binary system. One of the authors is queer, which means they have lived experience of the negative effects of the gender binary system on the LGBTQIA+ community. The authors span an age range of 30 years and thus have personal insights into how the gender system has changed over recent decades and affected people of different ages differently. Two of the authors are first generation highly educated, and one is a first generation full-time working mother.

Footnotes

[1] "trans*" is a term which is used to include people who may or may not identify as "transgender" but have an identity different from their assigned sex at birth. This includes those who fall under the non-binary umbrella.

[2] "Sommigen zijn wel erg (over)gevoelig wanneer ze niet op voorhand bediend worden van de meest ultieme uiting van politiek-correct taalgebruik die mogelijk is. [Bij 'dames en heren'] zit er voor 99,99% voor iedereen wat tussen. Voor de 0,01% hoeft niet iedereen zich [...] aan te passen."

[3] "Waar ik wat meer moeite mee heb, is dat [...] uitspraken [...] of m/v toiletten [...] omgezet moet[en] worden naar genderneutraal. Ik wil graag nog steeds aangesproken worden met dames of mevrouw. [...] Mijn identiteit hoeft niet te wijzigen hiervoor."

[4] "Ik vind [genderneutrale toiletten] ook behoorlijk naïef, want je blijft altijd viespeuken houden die de kans aangrijpen om vrouwen lastig te vallen."

[5] "Ik ben een geboren Amsterdammer van 74 jaar en herinner mij nog de simpele aanduiding van toen: hij of zij"

[6] "Groeten van een minderheid waarvan ik al 30 jaar niks hoor wegens het gejank van al dit soort aandachtstrekkers."

[7] Results regarding all differences between people's opinions depending on the article they read are beyond the scope of this paper but can be found in the Supplementary Materials.

[8] We also tested the validity of our measure, including correlations with previously validated items of experienced threat; the details of this can be found in the Supplementary Materials.

[9] For more details about our methodology and analyses, please see the Supplementary Materials.

Chapter 3: The Role of Gender Inclusive-Initiatives in Changing Beliefs about Gender and Stereotypes about Jobs

"Boys in the girl's room, girl's
in the men's room, you free
your mind in your androgyny"

Androgyny – Garbage

Author contributions: Miriam I. Wickham (Conceptualisation; Data curation; Formal analysis; Investigation; Methodology; Validation; Writing – original draft; Writing – review & editing); Félice van Nunspeet (Conceptualisation; Methodology; Validation; Writing – review & editing); Naomi Ellemers (Conceptualisation; Funding acquisition; Methodology; Validation; Writing – review & editing)

Abstract

Discussions about gender-inclusive initiatives (GII), which argue gender to be a spectrum, rather than binary (male and female), have been on the rise. These interventions are typically designed to increase inclusion of non-binary and transgender individuals. However, they may also benefit cisgender men and women by reducing gender stereotypes more generally, for instance, about which jobs are suitable for each gender. Across three studies, we explored whether a GII manipulation, as compared to a binary manipulation, may affect restrictive gender beliefs, stereotypes about gendered jobs, and interest in gendered jobs. In Study 1, we found that a GII manipulation affected explicit gender binary beliefs and gender essentialist beliefs, but not implicit beliefs about gendered jobs. We also found preliminary evidence that GII can affect interest in a Health, Elementary Education, and Domestic work (HEED) job. In Study 2, we focused on the effects of a GII manipulation on interest ratings of different gendered jobs. We found no evidence that GII can affect interest in jobs typically dominated by one gender. In Study 3, we found that a GII manipulation, as compared to a binary manipulation, affected participants' perceived societal beliefs about gender (i.e., what participants thought others in society believe). We also found an effect of the GII manipulation on explicit stereotypes about female-dominated jobs, but no evidence of GII affecting interest ratings in jobs. We conclude that GII have the potential to affect people's restrictive beliefs about gender in general, as well as their stereotypes about certain jobs.

Discussions about *gender-inclusive initiatives* (GII), which argue gender to be a spectrum, rather than binary, have been on the rise throughout the world (e.g., gender-neutral bathrooms, Blumell et al., 2019; gender-neutral pronouns, Sendén, et al., 2015; gender-markers in human-computer interactions, Jaroszewski et al., 2018). These interventions are typically designed to increase inclusion of gender non-binary or non-conforming individuals, though they have the potential to be beneficial beyond this (e.g., Bovens & Marcoci, 2020). For instance, GII challenge the idea that (cis)men and (cis)women are fundamentally different, and therefore that traditional gender roles are still relevant. As such, these initiatives may affect people's gender ideology and reduce gender-based stereotypes.

Negative Impacts of Gender Binary Beliefs

The *gender binary* is the view that humans all fit into one of two categories, men and women, that these categories are biologically meaningful, and that this view is natural and unmalleable (Hyde et al., 2019). The effects of this view are pervasive to anyone who identifies outside of the binary, but also for all other LGBT+ individuals (Garelick et al., 2017; Prusaczyk & Hodson, 2020). Because of this, advocates are arguing for GII to be increasingly implemented across society (e.g., Sawyer et al., 2016). However, GII may also have a positive impact on cisgender, straight people by reducing the idea that men and women are fundamentally different, and thereby reducing gender biases.

The gender binary is a self-perpetuating belief (Saguy et al., 2021), which is upheld and simultaneously upholds gender biases (i.e., the idea that one gender group is superior to the other in a certain domain) and subsequent gender discrimination (i.e., the exclusion of one gender group from a certain domain). Rather than a *natural* tendency to want to place others into one of two rigid categories, researchers have argued that the gender binary is culturally *learnt*,

as evidence from developmental psychology suggests (summarised by Hyde et al., 2019). This would mean that gender binary beliefs are malleable, and breaking the self-perpetuating cycle may also have a positive impact on gender biases. In the present study, we set out to test whether GII may be a tool to break this cycle.

How GII May Have a Positive Impact

GII raise awareness of gender being a spectrum in terms of identity (there are more than two gender identities), expression (men wearing skirts are still men), personality (women can be agentic), for instance. GII come in various forms, including gender-neutral toilets (Gershenson, 2010) and degendered language (Hord, 2016). In the present paper, we focus on popular media (newspaper, documentaries, etc.) initiatives that raise awareness of gender as a spectrum. These types of media attention are on the rise (Richards et al., 2017). They typically cover topics related to transgender and non-binary identities, intersex conditions, gender as a socially and culturally constructed idea, and how gender has been differentially perceived throughout history.

GII are typically designed to reduce discrimination of non-binary and gender non-conforming people. The idea of GII seems to be to convince people that gender is not binary, as well as that gender is not based in biology. In other words, GII are supposed to reduce gender binary and gender essentialist beliefs (i.e., the idea that biological differences between sexes justify the idea that men and women are fundamentally different from one another; Skewes et al., 2018). Because GII are underexamined, it is unclear whether they achieve this, though Morgenroth et al., 2021, did not find evidence that exposure to a GII reduced gender binary or gender essentialist beliefs. However, if GII are convincing enough to change people's gender beliefs, they may also have a positive impact on gender biases.

One type of gender bias that GII may be able to address are gender stereotypes about jobs, since having binary beliefs has been theorized to be related to labelling and sorting concepts such as, clothes, bathrooms, jobs, environments, into *masculine* or *feminine* categories (Saguy et al., 2021). Currently, Science, Technology, Engineering and Mathematics (STEM) jobs as still seen as more typical for men, and Healthcare, Elementary Education and Domestic (HEED) jobs are still seen as more typical for women (Block et al., 2018; Smeding, 2012). Because gender-typicality of jobs is held in place by stereotypes (Fiske et al., 2002), people who may be implicitly interested in, and talented at, a certain job may not pursue it because they think it is not suitable for someone with their gender. Furthermore, people may judge others for pursuing a job that is not considered suitable for them because of their gender. GII may increase the range of jobs that are considered appropriate for different genders by reducing the gender-based labelling and sorting of jobs. Consequently, GII may increase people's interest in gender-incongruent jobs because people are no longer being held back by their stereotypes as much.

The Present Studies

Across three studies, we researched the potential impact of GII, compared to binary manipulations, on gender binary beliefs, gender essentialist beliefs, gender stereotypes about jobs, and interest in jobs that are typically done by men or women. We had different methods and different foci in each study. Specifically, in Study 1, we investigated whether GII might affect implicit as well as explicit gender beliefs, a special feature of this study. Furthermore, we explored whether GII might affect men's and women's interest in STEM or HEED jobs. In Study 2, we zoomed in on whether GII might affect men's and women's interest ratings in different jobs, and how gender essentialist beliefs may affect biases about jobs being gendered. In Study 3, we investigated

whether GII may affect perceived societal beliefs (i.e., what people think that *others* in society believe), a special focus of this study. We also measured gender essentialist beliefs. Lastly, we looked at whether GII may affect people's biases about jobs being gendered.

Study 1

In Study 1, we compared a GII manipulation to a binary manipulation and a control (no manipulation), to test whether the GII manipulation would positively affect explicit beliefs about gender being binary, explicit gender essentialist beliefs, implicit gender biases about jobs, and/or interest in a STEM or HEED job. We expected that participants who are subjected to a GII manipulation more often report believing that gender is a spectrum, as compared to participants subjected to a binary manipulation. We also expected that participants who are subjected to a GII manipulation report holding less gender essentialist beliefs than participants subjected to a binary manipulation. A special feature of Study 1 was that we measured participants' implicit gender biases about jobs using their reaction times and accuracy in an Implicit Relational Assessment Procedure (IRAP; Cartwright et al., 2017). We thought that participants subjected to a GII manipulation may have less implicit gender biases than participants subjected to a binary manipulation. Moreover, participants were shown one of two job vacancies, one for a STEM job and one for a HEED job, and asked to indicate their interest and perceived fit in the job. We thought that participants who are subjected to a GII manipulation may show more interest in a gender-incongruent job vacancy than participants subjected to a binary manipulation. We expected to find no differences in any of the outcome measures between the binary and control condition, as society is already set up in a binary way and is therefore the default. Lastly, we also measured gender identification, to see whether that was a confounding variable because research has shown

that people with higher gender identification are more resistant to changes to the gender binary system (Morgenroth et al., 2021).

Method

Participants

Based on an effect size of g = 0.46 of the relationship between gender stereotype interventions in the form of heterogeneity (Lenton et al., 2009) and a 3 (conditions: binary, non-binary, control) x 2 (job gender-dominance: masculine, feminine) between-subjects design, we needed to collect 252 participants to achieve 80% power. We collected 255 participants from the online platform Prolific, all residing in the United Kingdom, with a mean age of 34.45 (SD = 12.62). Our participants' assigned sex at birth (ASAB) was evenly distributed, with 128 participants having been assigned male at birth (AMAB) and 127 having been assigned female at birth (AFAB). Participants' employment situation varied with a majority (35.39%) being employed full time. Similarly, participants' education level varied with a majority (32.94%) of participants having achieved a Bachelor degree. For detailed information about employment and education level, please see the Supplementary Materials.

Materials

Manipulation. Participants were shown one of three articles to either a) remind them of the binary system in society (there are only two genders: male and female), b) to explain that gender is not binary (GII manipulation), or c) a control article. All articles were extracted from a UK newspaper (The Guardian, https://www.theguardian.com/) and discussed findings from the field of Neuroscience. The binary article was about differences in male and female brains, the GII article was about was about similarities in male and female brains, and the control article was about neuroscience without mention of gender (see Appendices A, B and C respectively).

Implicit gender binary beliefs. To measure participants' implicit associations between gender groups and job type, we asked participants to do an Implicit Relational Assessment Procedure (IRAP) based on Cartwright et al. (2017). Participants were shown job words that are either typically feminine or typically masculine (e.g. nurses, teachers, scientists, engineers). Above each job word they either saw the words "men are" or the words "women are". Participants were asked to respond to each trial with "true" (pressing the letter 'q' on their keyboard) or "false" (pressing the letter 'p') according to a rule. If they responded incorrectly, they were shown a red 'X' for 500ms after the trial. The rule changed from block to block, such that there was one gender-congruent and one gender-incongruent block, as well as two practice blocks. In the gender-congruent block, the rule was that women are [typically feminine job] and men are [typically masculine job], and vice versa in gender-incongruent blocks. The order of the blocks as well as the order of the job words were randomised. Each block contained 32 trials. Participants were given a maximum of 2000ms to respond to each trial. After each trial, there was a break of 500ms before the next trial began. After each block, participants could rest for as long as they wanted before beginning the next block. By measuring the reaction times and accuracy of participants, we were able to calculate so-called d scores (Greenwald et al., 2003) to get an indication of how easy or difficult it is for participants to learn gender-congruent as compared to gender-incongruent rules. The IRAP makes it possible to extract four d scores, for the implicit beliefs that 1) women do typically feminine jobs, 2) men to typically masculine jobs, 3) women do not do typically masculine jobs, and 4) men do not do typically feminine jobs.[1]

Job Vacancies. Participants were shown one of two job vacancies, either a typically feminine (HEED) job or for a typically masculine (STEM) job. These were identical except for the title:

"Kindergarten Teacher Traineeship" and "IT Instructor Traineeship", respectively, and a few words pertaining to whom the applicant would teach (preschool children or employees at a company, respectively; see Appendices D and E). These job vacancies had been piloted among a different set of 21 Prolific participants (10 of whom viewed the feminine vacancy, 11 viewed the masculine vacancy), for how masculine and feminine the jobs were seen to be. In this pilot, we found that, although the vacancies only differed in terms of a few words, the STEM vacancy was perceived as more masculine than the HEED vacancy (M = 4.86, SD = 0.79; M = 3.90, SD = 1.42; $t(19) = -1.94$, $p = .067$; not significant likely due to sample size), and the HEED vacancy was perceived as more feminine than the STEM vacancy (M = 4.70, SD = 0.83; M = 4.07, SD = 0.51; $t(19) = 2.12$, $p = .048$).

Self-report. All items were measured on a 7-point Likert scale going from "Completely disagree" (1) to "Completely agree" (7).

Manipulation Checks. To measure how convincing our manipulations were, we asked participants the following questions about the article they read: "I believed that the article was real", "I believed the science that was referenced in the article", "I agreed with the sentiment of this article".

Gender Identification. To control for the possible confounding effect of gender identification on beliefs about the binary, whereby people who are high identifiers of their gender often hold more binary beliefs (Morgenroth et al., 2021; Wickham et al., 2021), we included four items. Two items were about gender identification with men, and two items about gender identification with women. The items were 'I identify with (other) men', 'I feel that I belong to the group of men' (α = .92) and the same two items with the word 'men' replaced with the word 'women' (α = .90).

Gender Binary Beliefs. To measure participants' explicit binary beliefs, we included a question about belief in 'gender-as-binary' or 'gender-as-spectrum', which was taken from a U.S. national poll (Fusion.net, 2015). "Some countries, including India, recognize a third gender that is neither male nor female. Which more closely aligns with your view?". There were three answer options: (a) "There are only two genders, male and female", (b) "Gender is a spectrum, and some people fall outside of binary categories" and (c) "I don't know".

Gender Essentialism. To measure participants' belief that gender is a biologically essentialist trait, we administered the gender essentialism scale from Skewes et al. (2018) which includes items such as 'Members of each gender have many things in common' (reverse coded) and 'Differences between women and men's personality are in their DNA' ($\alpha = .94$).

Perceived Fit of and Interest in Job. To measure how interested participants were in one of two job vacancies and how much they felt they would belong in a job (perceived fit), five items were written based on Tellhed et al. (2017), e.g., 'I would be interested in this job' and 'I would belong in this job'. We found these items to load onto the same factor (loadings all $\geq .90$) and reliability to be high ($\alpha = .95$).

Meeting of Job Requirements. The item 'I could fit the requirements of this job' was written in order to control for subjective applicability of the participant to one of two vacancies. This item did not load onto the same factor as items about perceived fit of and interest in job.

Procedure

We received ethical approval from the ethics board of our local Faculty of Social and Behavioural Sciences, which abides by APA ethical guidelines and the Declaration of Helsinki. Participants were recruited using the online platform Prolific. All participants gave

informed consent. They were redirected from Prolific to the survey and asked to answer all questions and do their best in the reaction time task. Participants saw materials in the following order: the (binary, non-binary or control) article, the (feminine or masculine) job vacancy, questions about belonging, interest in and requirements of the job, the IRAP, a control question about their dominant hand, belief in the gender binary question, gender essentialism scale, gender identification items, questions regarding their understanding of the article (manipulation checks) and lastly, demographic questions. For each questionnaire, the order of the items was randomized. They were shown a debriefing and given the opportunity to retract their data. Participants received payment for their participation.

Results

Manipulation Checks

To check whether our manipulations were convincing (how credible it was and how much participants agreed with the message of the article) we tested whether answers to manipulation checks differed significantly between conditions and whether the means were significantly higher than scale mid-point, using t-tests. Means of all manipulation check questions were significantly higher than scale mid-point and did not differ significantly per article condition (*ps* > .22). Participants believed the article to be real (M = 4.84, SD = 1.69, $t(254)$ = 7.87, $p < .001$, d = 1.69), believed the science referenced (M = 4.89, SD = 1.60, $t(254)$ = 8.92, $p < .001$, d = 1.60), and agreed with the sentiment of the article (M = 4.73, SD = 1.50, $t(254)$ = 1.87, $p < .001$, d = 1.50). This suggests that all our articles were convincing.

Gender Identification

Men's mean gender identification with men (M = 5.84, SD = 1.44) and women's mean gender identification with women (M = 5.93,

SD = 1.25) did not differ significantly between conditions (all *ps* ≥ .070). As such, we concluded that any differences between experimental conditions could not be attributed to differences in participants' gender identification.

Explicit Gender Binary Beliefs

In order to examine the effect of manipulation on explicit gender binary beliefs, we ran a one-sided chi-square test, including post-hoc z scores. This revealed (X^2 = 8.80, p = .033, φ = 0.19), in line with our expectations, that participants in the binary condition more often believed gender to be binary (52.32%) than a spectrum (43.02%), while participants in the GII condition more often believed gender to be a spectrum (63.53%) than binary (31.76%, post-hoc z's all p < .05). Participants in the control condition were somewhere in between, and not significantly different from participants in the experimental conditions (p > .05), believing that gender is binary in 39.29% of the cases, and a spectrum in 58.33% of the cases (all other participants not included in the reported percentages selected the option "I don't know"; we do not include these because we had no predictions about indifference of opinion). These results suggest that the articles had an effect on the binary beliefs of participants, in the expected direction.

Gender Essentialism

To investigate whether there was a main effect of manipulation on gender essentialism beliefs, we ran an ANOVA. This revealed that participants' mean gender essentialism beliefs (M = 3.57, SD = 1.03) did not differ significantly between experimental conditions (p = .167). This suggests that the content of the article did not affect participants' beliefs of gender being biologically essentialised, which is not in line with what we expected. However, given the effect of our manipulation on explicit gender binary beliefs, we wanted to examine how our

manipulated variable (explicit binary beliefs) may have further influenced gender essentialist beliefs.

To investigate whether mean gender essentialism differed between participants who believed gender to be binary and participants who believed gender to be a spectrum, we ran an independent samples t-test (IV: explicit binary beliefs, 2 levels: "gender is binary", "gender is a spectrum"). This revealed a significant difference (t (243) = 4.89, p < .001, d = 0.64), such that participants who believed gender to be binary held more gender essentialist beliefs (M = 3.91, SD = 0.91) and participants who believed gender to be a spectrum held less gender essentialist beliefs (M = 3.28, SD = 1.04). This suggests that explicit gender binary beliefs, which were affected by our manipulation, influenced gender essentialist beliefs.

Implicit Binary Beliefs about Jobs

To check for differences in implicit binary beliefs regarding the types of jobs men and women do, depending on condition (2 levels: binary and GII), we ran a MANOVA (4 DVs: d scores from the IRAP). We found no significant differences (ps ≥ .400) suggesting no effect of manipulation on implicit binary beliefs about jobs. We also checked whether explicit binary beliefs (2 levels: "gender is binary", "gender is a spectrum") further influenced implicit beliefs with a MANOVA (4 DVs: d scores) but found no significant results (ps ≥ .269). This suggests that there was also no further influence of the manipulated variable on implicit binary beliefs about jobs.

Relationship between Manipulation and Interest Ratings of Jobs

To test the idea that participants would show greater interest in a gender-incongruent job in the GII condition than in the binary condition, we ran two MANOVAs, one for each job (IT and kindergarten). With these, we tested whether the manipulation (2 levels: binary,

GII) had a significant effect on the fit and interest ratings of each job, or how much participants felt they could meet the requirements of each job, and whether this was moderated by participants' ASAB (2 levels: AMAB, AFAB). We found a main effect of manipulation on the fit and interest rating of the kindergarten job ($F(2, 121) = 4.81$, $p = .010$, $\eta_p^2 = 0.07$), but not on the belief that participants could meet the requirements, nor any main or interaction effect of ASAB ($ps \geq .068$). To understand the significant effect of manipulation on fit and interest ratings of the kindergarten job, we visually explored the descriptive statistics (see Table 1). The descriptive statistics suggested that participants in the GII condition, regardless of participants' sex, reported lower perceived fit and interest in the kindergarten job, than participants in all other conditions. However, they did not report thinking that they would meet the requirements of this job any less, suggesting that the effect of our article on perceived fit and interest was not driven by a lack of professional fit.

Table 1

Descriptive Statistics of Perceived Fit and Interest, and Meeting Requirements of each Job, per Condition

Condition	Job	Mean perceived fit and interest (*SD*)	Mean meeting requirements (*SD*)
Binary	IT	4.00 (1.71)	5.62 (1.64)
	Kindergarten	3.96 (1.59)	5.25 (1.70)
GII	IT	4.12 (1.65)	5.37 (1.81)
	Kindergarten	3.22 (1.47)	5.61 (1.63)
Control	IT	4.25 (1.56)	5.58 (1.65)
	Kindergarten	4.33 (1.61)	5.92 (1.16)

To further explore the significant effect of the GII manipulation on perceived fit and interest of the kindergarten job, we ran a planned contrast ANOVA with manipulation and job type as independent variables and perceived fit and interest as dependent variable. The contrast coefficients, based on our descriptive statistics, were as follows: binary*IT = 1, GII*IT = 1, control*IT = 1, binary*kindergarten = 1, GII*kindergarten = -5, control*kindergarten = 1. We chose these contrast coefficients because we wanted to test the data-driven idea that perceived interest and fit in the kindergarten job was lower in the GII condition than in any other condition, but that all other conditions did not significantly differ from each other. The contrast test was significant; $t(249) = 3.15$, $p = .002$, $d = 1.61$. Our results therefore suggest that the

GII article may affect participants interest in a job as a kindergarten teacher, but not interest in a job as an IT teacher. Our results also suggest that the binary manipulation and the control condition did not affect perceived fit and interest in either job.

Relationship between Explicit Binary Beliefs and Interest Ratings of Jobs

To examine whether our manipulated variable (explicit binary beliefs) further affected participants' fit and interest ratings of each job, or the belief that participants could meet the requirements of the job, we ran two MANOVAs, one for each job. With these, we tested whether participants' explicit binary beliefs (2 levels: "gender is binary", ": gender is a spectrum") affected the fit and interest ratings of each job, or how much participants felt they could meet the requirements of each job, and whether this was moderated by participants' ASAB (2 levels: AMAB, AFAB). We found no significant main or interaction effects (*ps* ≥ .124). This suggests that explicit binary beliefs did not further affect how participants felt about each job.

Discussion

We had two goals for our first study, namely, to investigate the effects of a GII manipulation, as compared to a binary manipulation and a control condition, on (1) explicit binary beliefs about gender and gender essentialist beliefs, and (2) on implicit gender binary beliefs about jobs and interest in gendered jobs. We found some evidence for our first goal, and limited evidence for the second goal. Taken together, we think that our gender-inclusive manipulation affected participants' beliefs about gender being binary, but not the (implicit) stereotypicality of jobs. Lastly, the GII manipulation may have had a small effect on participants' interest in a job that is typically done by women. We describe each result, in turn, below.

For our first goal, we had expected that participants subjected to a GII manipulation would more often report believing that gender is a spectrum, as compared to participants subjected to a binary manipulation or control condition, which we found. We had also argued that participants subjected to a GII manipulation may report having lower gender essentialist beliefs, as compared to participants subjected to a binary manipulation. We did not find a main effect of manipulation on gender essentialist beliefs. However, given that participants' explicit binary beliefs had been affected by the manipulation, we explored whether explicit binary beliefs further affected gender essentialist beliefs. Indeed, we found that participants who believed gender to be binary (who were more often in the binary condition) held more gender essentialist beliefs than participants who believed gender to be a spectrum (who were more often in the GII condition).

For our second goal, we had argued that participants who were subjected to a GII manipulation would hold less implicit stereotypical beliefs about jobs, than participants subjected to a binary manipulation. We found no effects of our manipulation on the d scores in the implicit measure, and we also did not find that our manipulated variable, explicit binary beliefs, further affected *d* scores. Contrary to our expectations, this suggests that our manipulation was not strong enough to change implicit beliefs about the stereotypicality of jobs. We think this might have been because implicit associations are harder to change than one's explicit beliefs (Gregg et al., 2006), as also shown by our results which revealed a significant effect of our manipulation on explicit binary beliefs, but not on the implicit binary beliefs about jobs. As such, in follow-up studies we wanted to measure explicit rather than implicit beliefs about job gender typicality. Furthermore, we had also argued that participants subjected to a GII manipulation may report greater interest in a gender-incongruent job vacancy than participants

subjected to a binary manipulation. We did not find evidence for this, nor did we find evidence for explicit binary beliefs affecting interest in the jobs. However, we did find an effect of our manipulation on the interest rating of the kindergarten job, regardless of participant ASAB. Specifically, only for the kindergarten job (and not the IT job), we found a difference in interest between conditions, such that participants in the GII condition were less interested than in the other two conditions. We thought this may be to do with perceived job status, as a kindergarten job may be seen as lower status than an IT job (Cuddy et al., 2008). A potential explanation, therefore, was that the GII manipulation made participants feel like they could opt out of the lower status job, regardless of whether the job was gender-congruent or not. However, because we did not measure the perceived status of each job vacancy in this study, we could only speculate. To investigate this further, in Study 2, we included more jobs of different status and measured the perceived status of each job.

We had expected no difference between the binary manipulation and the control condition in any test, as the gender binary is the default in society. In all our tests, the control condition did not significantly differ from either manipulation condition. However, the fact that it also did not differ from the GII manipulation is somewhat curious. We had expected the binary manipulation to represent the default of society and therefore for participants to behave the same way in the control condition as in the binary condition. However, the control condition was found to be "in between" the binary and the GII condition, suggesting that both manipulations differed from the default. We thought this could be because, though the gender binary is still a default in society, explicitly making participants aware of the gender binary may strengthen the effect of it, as compared to control. Considering that the control condition did not differ from either manipulation, however, we did not

see the added value of keeping a control condition in subsequent studies, as an extra condition would also affect the required sample size. In follow-up studies we therefore did not include a control condition.

In conclusion, we were successful in our first aim, to find effects of a GII manipulation, as compared to a binary manipulation, on gender binary and gender essentialist beliefs. However, we found limited evidence for our second aim, to find effects of a GII manipulation on gender-stereotypical beliefs about and interest in gendered jobs. In Study 2, we wanted to focus more on the latter, and made our primary aim to further explore under which conditions a GII manipulation might affect interest in jobs of different perceived status and gender typicality. To do so, we made some changes to the design. First, because we thought our manipulation may not have been strong enough to find effects on interest ratings of jobs, we chose a different manipulation, in the form of a video rather than an article. Research has shown that videos are more engaging than articles (Singhal & Rogers, 2012). Furthermore, to get a sense of how participants' interest in different jobs would differ before and after a manipulation, and to increase power, we changed the design to have repeated measures. We also asked participants to report their interest in a variety of jobs of different status and gender typicality, rather than only including two jobs in our design. To do so, we extracted a list of jobs with a range of different average salaries, and a range of how often they are done by women or men, from the UK Office for National Statistics. We also asked participants to rate jobs' societal status and gender typicality themselves. Furthermore, to be able to make a direct comparison between jobs, and to force participants to think about their preferences, we asked them to sort the jobs from highest to lowest interest, rather than rating each job on a scale. Lastly, as a secondary aim, to explore the effect of gender essentialist beliefs on explicit stereotypical beliefs about jobs, we also included measures for these in Study 2.

Study 2

Because the findings in Study 1 regarding interest in different jobs warranted further investigation, in Study 2, we focused on the effects of a GII manipulation, compared to a binary manipulation, on interest ratings of jobs. Specifically, we wanted to explore under which conditions participants' interest in jobs might be affected by a GII manipulation. Specifically, the conditions we explored were perceived societal status of the job, and gender typicality of the job. Based on the findings of Study 1, we thought that participants' interest in higher status jobs, or jobs typically done by men, may increase when subjected to a GII manipulation, as compared to lower status jobs, or jobs typically done by women, and as compared to a binary manipulation. As a secondary aim, we also explored how gender essentialist beliefs may be related to gender biases about jobs, where these were measured explicitly, rather than implicitly (due to no effect having been found in Study 1). We thought that there may be a positive relationship between holding greater gender essentialist beliefs and rating jobs in a more stereotypical way.

Method

Participants

Based on Lenton's et al. (2009) finding that gender stereotype interventions tend to have an overall effect size of $g = 0.32$, and our two-condition within-participants design (binary versus GII), we recruited 208 participants from Prolific to achieve 80% power. Participants all resided in the United Kingdom and had a mean age of 34.84 ($SD = 13.33$). Participants were evenly distributed across ASAB (104 each). In terms of gender identity, 104 participants were men, 101 were women, and three participants were non-binary; 18 participants were transgender (their gender identity did not match

their ASAB). Most of our participants were White (87.98%). Participants' employment situation varied with a majority (42.79%) being employed full time. Similarly, participants' education level varied with a majority (34.13%) of participants having achieved a Bachelor's degree. For more information about participant race, employment and education, please see Supplementary Materials.

Materials

Manipulation. Participants were shown two video clips, each to 1) remind them of the binary system in society (there are only two genders: male and female), and 2) to explain that gender is not binary (GII manipulation). The binary video explained how sex is determined by chromosomes stemming from the DNA of the father; it was cut to 57 seconds in length, and was extracted from Ted-Ed (Reedy, 2020). The GII video explained that there is scientific evidence that gender identity, expression, sex, and sexual attraction are not binary, it was cut to 72 seconds in length and was extracted from Netflix (Nye, 2017).

Self-report. All items were measured on a 7-point Likert scale (1 = "Completely disagree", 7 = "Completely agree"), unless indicated otherwise.

Gender Essentialism. To test how gender essentialist beliefs may affect people's ratings of jobs, we included the same items for gender essentialism as in Study 1 (α = .91).

Gender Identification. To test whether level of gender identification was a confound for any results, we included the gender identification scale from Wickham et al. (2021). It includes items such as "Men are an important reflection of who I am" and "I see myself as someone belonging to the group of women" (items about women, α = .94; items about men, α = .91).

Own Status. In order to explore whether own status could affect results regarding jobs, we included an item for subjective social status (Goodman et al., 2001). Participants place themselves on a 10-rung ladder (answer options 1-10, higher numbers meaning higher status), based on money, schooling, and the respect that society gives them.

Attention Checks. Participants were asked three multiple-choice questions about each video, to check that they had watched and understood the content of the videos. Furthermore, we monitored how much time participants spent on each page of the survey, and how many clicks they made in the sorting tasks. We excluded participants who failed two or more attention checks (questions answered incorrectly, less time spent on video page than duration of video, less than three clicks during one sorting task).

Jobs. We extracted job titles from the UK Office for National Statistics, and included 16 jobs from a range of sectors, with a range of wage estimates (an indicator of job status) and gender ratios (an indicator of gender-dominance). The jobs selected were all done by at least 50,000 people to ensure that these jobs were relatively common and known. The name of the job was occasionally changed to make sure they were relatively common words, e.g., "Cleaner" or "Manager". Each job title was given a description; all descriptions were extracted and edited from Workable (resources.workable.com; Appendix F).

Sorting Tasks. Participants were asked to sort the 16 jobs according to gender typicality, personal interest, likelihood of having the jobs themselves, and perceived status. They did this by dragging and dropping the job titles until they were in the order the participant wanted. Sorting position 1 was at the top, while position 16 was at the bottom, meaning the lower the number (1-16), the higher the participant's rating.

Gender-typicality Sorting. Participants were asked to consider that some jobs are done more often by men, and some jobs are done more often by women. They then sorted the jobs from most typical for women (1) to least typical for women (16). Next, they sorted the jobs from most typical for men (1) to least typical for men (16).

Interest Sorting. Participants were asked to imagine that they wanted to choose a new career and to disregard what their CV may be suitable for. They then sorted the jobs from most interesting (1) to least interesting (16).

Likelihood Sorting. Participants were asked to consider how likely it was that they would ever have this job in real life, given their education and skills. They then sorted the jobs from most likely (1) to least likely (16).

Status Sorting. Participants were asked to consider the status that people with each job have in society, based on money, schooling and respect (Goodman et al., 2001). They then sorted the jobs from best off (1) to worst off (16).

Filler Sorting Tasks. To distract participants from the purpose of the experiment, we included three additional sorting tasks, where participants sorted European countries, hobbies and dessert according to personal preference.

Procedure

Ethical approval, recruitment and informed consent procedures were the same as in Study 1. Participants were asked to answer all questions according to their gut feeling (except for attention checks). They first saw the binary video and answered the manipulation check questions. Next, they viewed the job descriptions, sorted the jobs according to interest, and completed the filler sorting tasks. Next, they viewed the GII video, answered manipulation check questions, and sorted

the jobs according to interest for the second time. Next, they completed the likelihood, status, and gender-dominance sorting tasks. Lastly, they filled in the gender identification, gender essentialism, own status, and demographic items. For each questionnaire, the order of the items was randomized. They were shown a debriefing and given the opportunity to retract their data. Participants received payment for their participation.

Results

Self-report Measures

For descriptive statistics of self-report items, please see Table 2.

Table 2

Descriptive Statistics of Gender Identification, Gender Essentialism and Perceived Own Status

Measure	M (*SD*)
Women's gender identification with women	5.13 (1.23)
Men's gender identification with men	4.56 (1.22)
Gender essentialism	3.87 (0.88)
Own status (scale 1-10)	5.30 (1.83)

Status and Gender Typicality Sorting of Jobs

For mean status and gender typicality ratings per job, please see Table 3. Because we live in a patriarchal society where men tend to have higher status than women, we expected that jobs which are typical for men would be seen as having higher status than jobs that are typical for women. To investigate this, we ran a Pearson's correlation analysis. We found a significant negative correlation between typicality for women and status rating (r = -.05, p = .004) and a significant positive correlation between typicality for men and status rating (r = .25, p < .001), as expected. However, considering the effect sizes of these correlations being relatively small, we visually explored our descriptive statistics, and thought that we may still had a good spread of jobs of different status and gender typicality for our purposes. Furthermore, on the whole, our findings of job status and gender typicality fit with what we had expected based on the data of the UK office for National Statistics.

Table 3

Descriptive Statistics of Status and Gender Typicality Ratings

Job	Status rating, *M* (*SD*)	Typicality for women, *M* (*SD*)	Typicality for men, *M* (*SD*)
Doctor[m]	1.49 (1.60)	10.88 (3.81)	3.95 (2.63)
Manager[m]	3.16 (1.99)	10.82 (3.25)	4.67 (2.76)
Police officer[m]	4.37 (2.42)	12.14 (2.88)	3.87 (2.64)
Fashion consultant[f]	5.13 (3.13)	5.38 (3.36)	12.35 (2.99)
Nurse[f]	5.46 (2.91)	3.25 (2.52)	12.02 (3.18)
Social worker[f]	6.71 (3.16)	6.73 (3.34)	10.25 (2.96)
Administrative secretary[f]	7.44 (2.68)	5.86 (3.26)	11.24 (2.85)
Travel agent[n]	8.02 (2.54)	8.43 (3.04)	9.23 (2.88)
Receptionist[f]	9.73 (2.49)	4.95 (2.97)	11.68 (2.89)
Postal worker[m]	9.99 (2.45)	12.18 (2.45)	5.72 (2.87)
Truck driver[m]	10.12 (3.28)	14.93 (2.16)	2.36 (2.41)
Call centre representative[n]	11.58 (2.57)	9.77 (3.09)	8.26 (3.13)
Cashier[n]	12.06 (2.76)	8.01 (3.37)	9.52 (2.87)

Factory production worker[m]	12.23 (2.95)	12.95 (2.64)	5.00 (3.15)
Babysitter[f]	13.95 (2.65)	3.70 (3.02)	13.94 (2.90)
Cleaner[f]	14.55 (2.19)	6.02 (3.48)	11.95 (3.23)

Note. Jobs were sorted from 1 to 16, with lower numbers signifying higher status and gender typicality. Jobs are shown in order from highest to lowest mean status. Jobs are denoted with [f] if they were perceived to be typical for women, denoted with [n] if they were perceived to be relatively neutral, and denoted with [m] if they were perceived to be typical for men.

Based on the mean status rating scores we classified jobs as high status (Manager, Doctor, Police officer, Fashion consultant: $M < 5.33$), medium status (Nurse, Postal worker, Administrative secretary, Truck driver, Social worker, Travel agent, Receptionist; $5.33 < M < 10.67$), and low status (Factory production worker, Call centre representative, Babysitter, Cashier, Cleaner; $M > 10.67$). Based on the difference between the mean female typicality score and the mean male typicality score (*diff*) we classified jobs as typically masculine (Manager, Doctor, Police officer, Postal worker, Truck driver, Factory production worker; *diff* > 1.51), neutral (Call centre representative, Travel agent, Cashier; $-1.51 < diff < 1.51$), or typically feminine (Fashion consultant, Nurse, Administrative secretary, Social worker, Receptionist, Babysitter, Cleaner; *diff* < -1.51). Thus, though jobs which were seen as typical for men were often also seen as higher status, we did find various jobs of different gender typicality and status. As such, we concluded that we had succeeded in getting a good spread of jobs in terms of status and gender, for the purposes of our follow-up analyses.

Relationship between Manipulation and Interest in Job, Depending on Job Status and Gender Typicality

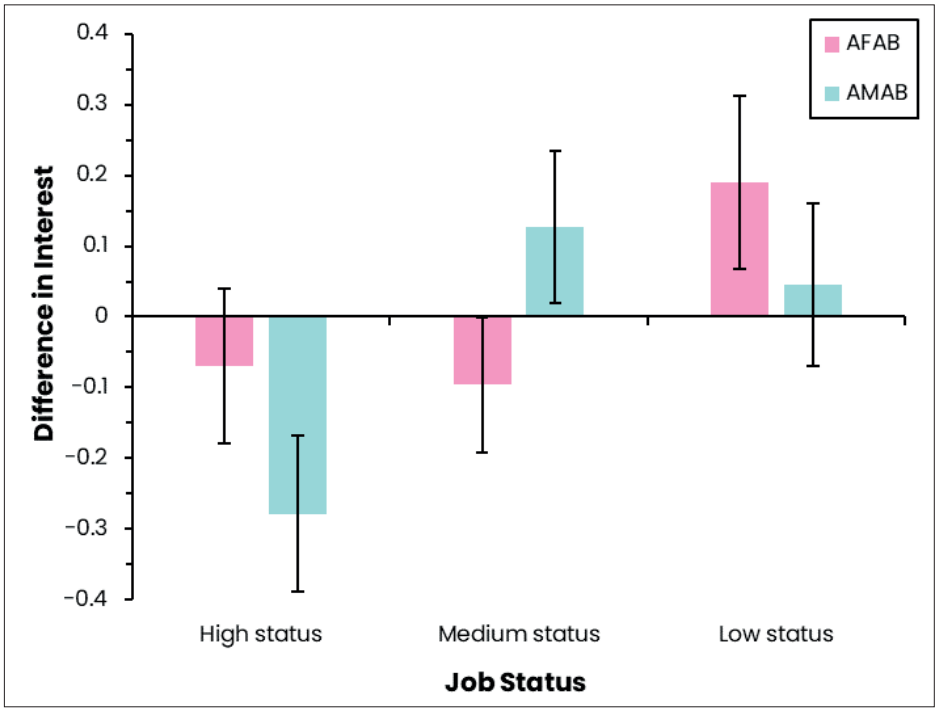
To explore whether participants' interest changed from time 1 (binary condition) to time 2 (GII condition) depending on job status and gender typicality, we ran two ordinal regressions (one for AMAB participants and one for AFAB participants) with job status and job gender typicality as independent variables and the difference score (time 2 – time 1) in interest ratings as dependent variable. This revealed no significant effects of job status or gender typicality on interest change from time 1 to time 2 (all $ps \geq .152$). This suggests no effect of our manipulation, contrary to our expectations.

As can be seen in Figure 1, the pattern of results (albeit non-significant) was also in the opposite direction than we expected based

on Study 1, with participants' interest in low status jobs generally increasing from time 1 to time 2, and participants' interest in high status jobs generally decreasing from time 1 to time 2. Though not a significant result, this suggests that the GII manipulation (time 2) might have increased interest in low status jobs and decreased interest in high status jobs, which is the opposite of what we had thought.

Figure 1

Differences in Interest (Time 2 – Time 1) Depending on Job Status and ASAB



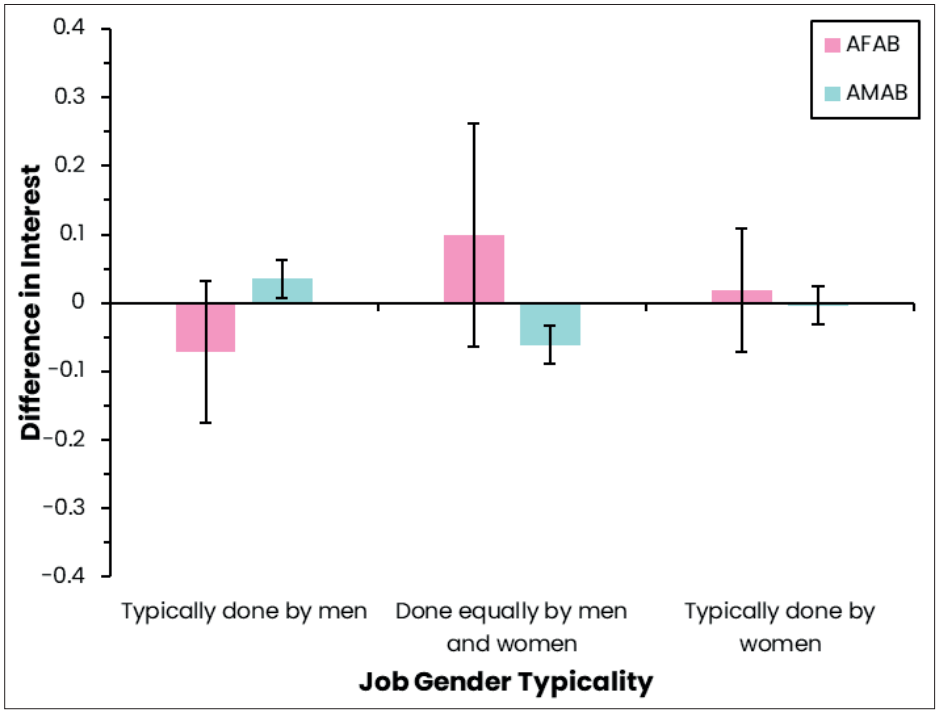
Note. Error bars represent the standard error of the mean.

Furthermore, as can be seen in Figure 2, female participants also seemed to gain interest in jobs typical for women at time 2, and male participants seemed to gain interest in jobs typical for men at time 2, compared to time 1. Though not a significant result, this suggests, that

the GII manipulation (time 2) might have increased interest in gender-congruent jobs, contrary to what we had thought.

Figure 2

Differences in Interest (Time 2 – Time 1) Depending on Job Gender Typicality and ASAB



Note. Error bars represent the standard error of the mean.

Effect of Gender Essentialism on Gender Typicality Rating

Since in Study 1 we had not found an effect of gender essentialism on *implicit* binary beliefs about jobs, in Study 2 we wanted to explore whether gender essentialism might be related to *explicit* binary beliefs about jobs (i.e., gender typicality ratings). We thought that gender essentialist beliefs might affect gender typicality ratings of jobs, such that participants with higher gender essentialist beliefs might rate typically male jobs as *more* typically male, and typically female jobs as

more typically female, as compared to participants with lower gender essentialist beliefs. As such, we ran a linear model with gender typicality classification (a categorical variable constructed based on the gender typicality mean scores of the whole sample) as independent variable, gender essentialism as moderator, and mean gender typicality rating (a continuous variable with individual variation) as dependent variable. In other words, we wanted to see whether, in the relationship between group classifications of jobs (in terms of gender) and individual scores of jobs (in terms of gender), gender essentialism was a moderator. However, we found no significant effect of gender essentialism (p = 1.000).

Discussion

We had two aims for Study 2: (1) to investigate whether participants' interest in jobs of different perceived societal status and gender typicality would be affected by a GII manipulation, as compared to a binary manipulation, and (2) to explore the relationship between gender essentialist beliefs and gender typicality ratings of jobs. We did not find conclusive evidence for either of our research questions. We describe each of our inconclusive findings, in turn, below.

In Study 1 we had found, in an exploratory and data-driven test, that participants' interest in a job typically done by women, and which is typically low status, was lower in the GII condition than the binary condition. Based on this result, we wanted to investigate further the potential effects of a GII manipulation on participants' interest in jobs of different status and gender typicality. We had thought, based on our previous results, that participants would show greater interest in high status jobs, or jobs typically done by men, after having been shown the GII manipulation (time 2), as compared to after the binary manipulation (time 1). As seen by the non-significant result of the ordinal regressions, we did not find evidence for this. In fact, though this was a non-significant result, visual exploration of our data suggested

that participants' might have gained interest in lower status jobs after the GII manipulation, and that women might have gained interest in jobs typically done by women, while men gained interest in jobs typically done by men. All participants, in turn, seemed to have lost interest in high status jobs after the GII manipulation. This is in the opposite direction of what we had anticipated.

The reason for the non-significant results for interest may have been because someone's interest in a job is not as easily malleable (perhaps it is more of a trait than a state; Steyer et al., 1999) as other factors, such as their beliefs about gender being binary, for which we had found an effect in Study 1. Furthermore, though we had thought that a within-subjects design might give us insight into how participants' interest ratings of jobs differed before and after a manipulation, and although we had included distractor questions in between ratings, there may still have been a response bias. Specifically, participants may have tried to rank jobs similarly before and after the manipulation. We had also tried to circumvent response bias, especially in the form of neutral responding (i.e., bias towards the scale mid-point), by asking participants to rank jobs rather than rate them on a scale. We thought this would force them to think about their choices more. However, this also meant that we lost some of the nuances of the data, as participants always compared interest in a particular job to interest in another job. Had we used a Likert scale to measure interest in a particular job before and after, we would have been able to measure smaller differences in interest due to manipulation. So, though we had good reasons to try a within-participants design with rank order data, we decided to investigate whether we could clarify the non-significant results of Study 2, or find different results with a between-participants design and continuous data in Study 3.

Our second aim was to explore the relationship between gender essentialist beliefs and gender typicality ratings of jobs. We had thought

that participants who hold highly gender essentialized beliefs would consider jobs to be more stereotypically gendered than participants who hold less gender essentialized beliefs. This is because gender essentialism is the belief that gender is rooted in meaningful, biological differences between the sexes, and this may be related to the idea that feminine jobs are more suited to women because of their nature, and that masculine jobs might be more suited to men because of their nature. However, we did not find evidence for this exploratory idea. However, investigating the relationship between gender essentialist beliefs and gender typicality ratings was not our primary focus, and we had designed our study with the aim of finding differences in interest ratings of jobs in mind. As such, a limitation to this study was that our design prevented us from being able to explore how gender essentialist beliefs, and gender typicality ratings, might have been affected by our manipulation. Instead, we could only explore how these two measures were related to each other, because both measures were measured at the end of the study (after time 2). We therefore decided to see whether we would find different results in a between-subjects design in Study 3.

In conclusion, we did not find conclusive evidence for either of our research questions in Study 2. In Study 3, our main aim was therefore to focus once more on the effects of a GII manipulation, as compared to a binary manipulation, on beliefs about gender, for which we had found interesting findings in Study 1. To do so, we included measures for beliefs about society (with regards to gender), and gender essentialist beliefs. Furthermore, we aimed to investigate whether the inconclusive findings of Study 2 might have been related to the design of the study. Specifically, we wanted to investigate whether participants gender-typicality ratings of jobs would differ depending on manipulation (GII or binary), where gender essentialism or other beliefs about gender could be moderators. Moreover, we

wanted to further explore whether we could find any effects on participants' interest ratings of jobs, depending on condition. As such, in Study 3, we made sure to use a between-subjects design and Likert scale measures and included measures for interest in a job, perceived societal status of a job, and gender-typicality of a job. Additionally, to be able to compare the results of Studies 1 and 2, we combined the GII manipulations and list of jobs. This allowed us to see whether the different manipulations and different job titles could have resulted in different findings between studies. Lastly, we included another confounding variable in Study 1 which could affect participants' interest ratings of jobs: an order effect of questions. Specifically, we thought that asking participants about perceived status of a job, before asking them about their interest in a job, might affect their interest ratings. This is because asking about status is a subtle reminder of which job is most desirable. To be able to account for such a confounding factor, participants were randomly assigned to one of two conditions (besides the manipulation conditions): one where participants were asked about status before interest, and one where participants were asked about interest before status.

Study 3

Since we found no effect on interest ratings of jobs in Study 2, and because of the interesting findings regarding gender binary and essentialist beliefs in Study 1, in Study 3 we switched our focus again. Specifically, in Study 3, we investigated the effects of a GII manipulation, as compared to a binary manipulation, on perceived societal beliefs (rather than personal beliefs) about gender being binary, gender essentialist beliefs, and explicit biases about jobs. We expected that participants subjected to a GII manipulation would report that society at large generally agreed that gender is not binary. On the other hand, participants subjected to a binary manipulation

would report that society at large generally agreed that gender is binary. We also thought that participants' gender essentialist beliefs would differ depending on manipulation, and/or depending on their perceived societal beliefs about gender. Furthermore, we expected that participants subjected to a GII manipulation would hold less explicit gender biases about jobs, than participants subjected to a binary manipulation. Lastly, we explored interest ratings of jobs to see whether would replicate our null effects from Study 2. A special feature of Study 3 was that we combined the manipulations and list of jobs from Studies 1 and 2, in order to be able to compare the materials from each study with each other.

Method

Participants

Based on an effect size of g = 0.46 of the relationship between gender stereotype interventions in the form of heterogeneity (Lenton et al., 2009), and a 2 (gender condition: binary, GII) x 2 (order condition: status first, interest first) between-subjects design, we needed N = 186 to reach a power of 80%. We recruited 201 participants from Prolific, mean age 28.38 (SD = 9.97), 100 of them were AFAB and 101 AMAB. In terms of gender identity, 100 participants were men, 99 participants were women, and two participants were non-binary; three participants were transgender (their gender identity did not match their ASAB). Participants were distributed among 21 countries, with most residing in the United Kingdom (29.69%). Most of our participants were White (86.57%). Participants' employment situation varied with a majority (35.82%) being employed full time. Similarly, participants' education level varied with a majority (25.87%) of participants having achieved a Bachelor degree. For more information regarding participants' country of residence, race, employment, and education, please see Supplementary Materials.

Materials

Manipulation. Participants either viewed the binary article from Study 1 and the binary video from Study 2, or the GII article from Study 1 and the GII video from Study 2. Whether the participant first viewed the article or the video was randomised.

Perceived Societal Beliefs about Gender. Furthermore, to gain insight into whether participants felt that the materials shown to them were convincing, and to see how the manipulations affected their beliefs about society, we asked them to rate, on a 7-point Likert scale (1 = "Completely disagree", 7 = "Completely agree"), how much most people in society would agree with various statements. Statements were constructed based on quotes from the articles/ videos themselves. An example statement about the GII video was "Gender is a spectrum and there are more categories than just 'male' and 'female'" (see Appendix G; four items about binary video, $\alpha = .68$; three items about binary article, $\alpha = .67$; four items about GII video, $\alpha = .78$; three items about GII article, $\alpha = .35$, these items were analysed separately due to insufficient reliability).

Gender Essentialism. We included the same self-report measure for gender essentialism as in Study 2 ($\alpha = .91$).

Gender Identification. We included the same measure

as in Study 2 (gender identification with women, $\alpha = .90$; gender identification with men, $\alpha = .91$).

Own Status. We included the same measure as in Study 2.

Jobs. We included the same jobs and job descriptions as in Study 2. We additionally added the jobs "Kindergarten teacher" and "IT instructor", from Study 1, and provided them with similar job descriptions, based on Workable (resources.workable.com; Appendix F).

Measures about Jobs. Instead of participants sorting or ranking jobs according to interest, status, and gender-dominance, participants rated the jobs on a scale (7-point Likert scale for interest; 1 = "Not interested at all", 7 = "Extremely interested"; and gender-dominance; 1 = "Mostly done by women", 7 = "Mostly done by men"; 10-point scale for status; 1 = "Low status", 10 = "High status"). Gender typicality questions were combined into a single item, instead of two, going from 1 ("Mostly done by women") to 7 ("Mostly done by men").

Attention Checks. Participants were asked one multiple choice question about each video and each article, to check that they had attended to and understood the materials. We also monitored how much time was spent on each video page. Lastly, we included two attention checks within the self-report surveys ("Please choose completely agree"). Participants who spent less time on the video page than the duration of the video, and/or answered two or more attention questions incorrectly, were rejected on Prolific.

Procedure

Ethical approval, recruitment and informed consent procedures were the same as in Study 1. Participants were asked to answer all questions according to their gut feeling (except for attention checks). First, participants viewed materials in the following order: manipulation, perceived societal beliefs items, job descriptions. Next, they were either asked to rate jobs according to interest and then status (order condition: interest first), or according to status and then interest (order condition: status first). Next, they rated jobs according to how much they are done by men and women (gender typicality rating), they filled in the gender identification, gender essentialism, and own status questions, and filled in their demographics. They were shown a debriefing and given the opportunity to retract their data. Participants received payment for their participation.

Results

Perceived Societal Beliefs

To be able to investigate whether our manipulations were convincing and affected participants' perceptions of societal beliefs regarding gender being binary and gender categories being meaningful, we tested whether our items differed significantly from scale mid-point. In a series of one-sample t-tests, we found that participants in the binary condition believed that most people in society would agree with the statements made about both the video (mean of all items significantly higher than scale mid-point; $M = 5.50$, $SD = 0.93$, $t(99) = 16.09$, $p < .001$, $d = 0.93$; see Table 4), as well as the article (mean of all items significantly higher than scale mid-point; $M = 5.35$, $SD = 1.08$, $t(99) = 12.49$, $p < .001$, $d = 1.08$; see Table 4). This suggests that participants in the binary condition thought that most people in society would agree that gender is binary.

Participants in the GII condition believed that most people in society would agree with the statements, regarding gender being a spectrum and individual differences in gender expression, made about the video (mean of all items significantly higher than scale mid-point in one-sample t-test; $M = 4.75$, $SD = 0.98$, $t(100) = 7.69$, $p < .001$, $d = .98$; see Table 4). They were less convinced by the statements made about the article. Participants' mean scores for two of the items, regarding differences in male and female brains and how acceptable it is to act outside of a gender role, were somewhat lower and not significantly different from scale mid-point ($ps \geq .115$). Participants did report thinking that most members of society would agree that some men have more 'female' brains and some women have more 'male' brains ($t(100) = 4.01$, $< .001$, $d = 1.39$; see Table 4). This disparity between article and video could account for some differences in results between Studies 1 and 2, as the article was used in Study 1 and the video in

Study 2. Nonetheless, since participants in the GII condition were shown both the video and the article, and participants did agree with the statements about the video, we concluded that participants in the GII condition had indeed been affected by the manipulation. Importantly, participants also did not disagree with the items about the GII article (i.e., scores were not significantly lower than scale mid-point). As such, we believe that our video and article were convincing enough to affect their binary beliefs to some extent.

Table 4

Descriptive Statistics of Items Regarding Beliefs about Society, which Suggest that Perceived Societal Beliefs were Affected by our Manipulations

Condition	Manipulation type	Perceived societal belief item	*M* (*SD*)
Binary	Video	"When someone is expecting a child, they are very interested in finding out whether it's a boy or a girl"	5.50 (0.93)*
		"Most women want to have biological children at some point"	
		"Men focus more on their careers, while women focus more on family life"	
		"Baby boys and baby girls are different, it's natural"	
	Article	"Male and female brains are wired differently"	4.75 (0.98)*
		"The biological differences between men and women can explain why men and women fulfil different roles in society"	
		"Women are better at doing social activities, while men are better at co-ordination"	
GII	Video	"Gender is a spectrum and there are more categories than just 'male' and 'female'"	5.35 (1.08)*

	"Sex is binary, and your hormones, chromosomes and genitals will always align with each other" (R)	
	"Individuals are all different and that's ok"	
	"Everyone should be able to choose how they present themselves to the world, for instance, in how they dress or how they act"	
Article	"There are no real differences between male and female brains"	3.73 (1.91)
	"Some men have more 'female' brains and some women have more 'male' brains"	4.55 (1.39)*
	"Gender roles are largely cultural and it's ok for men and women to act outside of their roles"	4.28 (1.75)

Note. Participants had been instructed to rate how much they thought most people in society would agree with each of the statements. The items about the binary video, binary article, and GII video were analysed as a scale, while the items about the GII article were analysed separately. Means denoted with an asterisk (*) significantly differed from scale mid-point.

Confounding Effects of Gender Identification and Own Status Ratings

To check for the possible confounding effect of gender identification with men, with women, or own status ratings on our results, we tested, in a MANOVA, whether these differed between conditions (2 levels: binary, GII) and found no significant differences (*ps* ≥ .284; see Table 5). This suggests that neither gender identification nor own status were an alternative explanation for our other results.

Table 5

Descriptive Statistics of Gender Identification and Own Status Rating

Measure	M (SD)
Women's gender identification with women	5.31 (1.02)
Men's gender identification with men	5.13 (0.99)
Own status rating	5.61 (1.57)

Note. Own status rating was measured on a 10-point scale, while the other items were rated on a 7-point Likert scale.

Effect of Manipulation on Gender Essentialist Beliefs

To investigate whether our manipulation (2 levels: binary, GII) affected gender essentialist beliefs, we ran an ANOVA which revealed no significant differences (p = .463) in gender essentialist beliefs between conditions. Participants' mean gender essentialism across both groups was 3.95 (SD = 0.90).

Following the methods of Study 1, where we found that the manipulated variable, explicit binary beliefs, further affected gender essentialist beliefs, we wanted to see whether perceived societal beliefs about gender would similarly affect gender essentialist beliefs in this study. We therefore ran two linear regressions (one per condition), with perceived societal beliefs (about video, about article) as independent variables, and gender essentialism as dependent variable. We found no

significant effects (*ps* ≥ .317) suggesting that the manipulated variable in this study (perceived societal beliefs) did not further affect gender essentialist beliefs.

Status Rating and Gender Typicality of Jobs

We wanted to see whether the status and gender typicality ratings of jobs were similar to what we had previously found in Study 2. For mean status and gender typicality ratings per job, see Table 6. These descriptive statistics gave us an indication for which jobs are seen as relatively high or low status, or more typical for men or for women. This was mostly in line with what we expected based on the results of Study 2, with some minor differences (e.g., "Nurse" being seen as relatively high rather than medium status, which could be explained by the outbreak of the covid-19 pandemic meaning those working in the healthcare system were appreciated more). Pearson's correlations revealed that, across both gender groups there was a significant positive relationship between status and gender typicality (r = .15, p < .001, for men; r = .13, p < .001, for women), suggesting that participants ascribed higher status to more masculine jobs, as one might expect based on gender stereotypes, and in line with results from Study 2.

Based on these descriptive statistics, we classified jobs into three status categories[2]: low status (Cleaner, Call center representative, Babysitter, Factory production worker, Cashier, Truck driver; M^{status} > 6.20), medium status (Receptionist, Postal worker, Travel agent, Fashion consultant, Administrative secretary, Social worker; 6.20 > M^{status} > 4.60), and high status (Kindergarten teacher, IT trainer, Police officer, Nurse, Manager, Doctor; M^{status} < 4.60). We also classified jobs into three gender typicality categories, based on difference from scale mid-point3: typically female (Babysitter, Kindergarten teacher, Nurse, Receptionist, Cleaner, Administrative secretary, Fashion consultant, Social worker; M^{gender} < 3.00), relatively neutral (Cashier, Travel agent, Call center

representative, Doctor, Postal worker, Manager; 5.00 > M^{gender} > 3.00), and typically male jobs (Factory production worker, IT trainer, Police officer, Truck driver; M^{gender} > 5.00).

Table 6

Mean Status and Gender Typicality Ratings per Job

Job	Status rating, *M* (*SD*)	Gender typicality rating, *M* (*SD*)
Doctor[n]	9.18 (1.15)	4.68 (0.91)
Manager[n]	7.57 (1.69)	4.93 (1.07)
Nurse[f]	7.32 (1.79)	2.04 (1.02)
Police officer[m]	7.07 (1.72)	5.63 (1.05)
IT trainer[m]	6.60 (1.52)	5.61 (1.08)
Kindergarten teacher[f]	6.29 (1.81)	1.74 (1.00)
Social worker[f]	6.13 (1.90)	2.91 (1.14)
Administrative secretary[f]	5.64 (1.63)	2.32 (1.07)
Fashion consultant[f]	5.56 (2.13)	2.63 (1.24)
Travel agent[n]	5.11 (1.62)	3.52 (0.99)
Postal worker[n]	4.82 (1.58)	4.82 (1.21)
Receptionist[f]	4.72 (1.48)	2.26 (1.09)
Truck driver[m]	4.52 (1.83)	6.51 (0.78)
Cashier[n]	4.22 (1.72)	3.17 (1.06)
Factory production worker[m]	4.18 (1.81)	5.30 (1.26)
Babysitter[f]	4.11 (1.91)	1.42 (0.75)
Call center representative[n]	4.01 (1.73)	3.67 (0.91)
Cleaner[f]	3.40 (1.97)	2.28 (1.22)

Note. Status was measured on a 10-point scale, while gender typicality was measured on a 7-point scale. Lower gender typicality ratings represent a job being more typical for women, while higher gender typicality ratings represent a job being more typical for men. Jobs are ordered from highest to lowest status. Jobs are denoted with [f] if they were

perceived to be typical for women, denoted with ^{n} if they were perceived to be relatively neutral, and denoted with ^{m} if they were perceived to be typical for men.

Predicting Gender Typicality Ratings

In previous studies we had not found an effect of manipulation on gender binary beliefs about jobs (implicit or explicit). Nonetheless, we wanted to explore whether our manipulation had an effect on job gender typicality ratings (i.e., explicit binary beliefs about jobs). Furthermore, since we had not found an effect of manipulation on gender essentialist beliefs, we wanted to explore whether gender essentialism was a moderator of the effect of manipulation on gender typicality ratings. We also wanted to investigate whether the effect of manipulation on gender typicality ratings differed depending on the interest or status ratings of the job (i.e., whether interest or status were moderators). We tested these research questions using a series of three general linear models (one for each gender typicality classification; adjusted alpha level to account for multiple tests using Bonferroni correction; $p < .017$ is thus considered significant). In all cases, the dependent variable was gender typicality rating, the independent variable was the manipulation (2 levels: binary, GII), and gender essentialist beliefs, interest rating of the job, and status rating of the job were moderators.

Jobs That Are Typical for Women. There was a significant main effect of manipulation on gender typicality rating ($F(1,1595) = 22.87$, $p < .001$, $\eta_p^2 = 0.01$, see also Table 7). This suggests that the GII manipulation affected participants' gender typicality ratings of typically feminine jobs, such that they rated them as more neutral (meaning: less binary, closer to scale mid-point) compared to participants in the binary condition. We also found gender essentialist beliefs to be a significant moderator of this effect ($F(1,1595) = 16.82$, $p < .001$, $\eta_p^2 = 0.02$), such that the effect of the binary manipulation was weaker for people low on gender essentialist

beliefs. Furthermore, we found interest ($F(1,1595) = 16.32$, $p < .001$, $\eta_p^2 = 0.01$) and status ($F(1,1595) = 5.74$, $p = .012$, $\eta_p^2 = 0.01$) ratings of the jobs to be significant moderators of the effect. Specifically, we found that jobs with higher interest ratings were rated in a less stereotypical (i.e., more neutral) way, and jobs that were rated as high status were rated in a less stereotypical (i.e., more neutral) way. Taken together, we concluded that the GII manipulation, compared to the binary manipulation, affected participants such that they viewed typical jobs for women as less typical for women, especially if they were interested in the job, considered the job to be high status, and had lower gender essentialist beliefs.

Jobs That Are Relatively Neutral. We found no effect of manipulation on gender typicality ratings for neutral jobs ($p = .414$).

Jobs That Are Typical for Men. We found no effect of manipulation on gender typicality ratings for jobs typical for men ($p = .382$).

Table 7

Descriptive Statistics of Gender Typicality Rating per Job Type, per Condition

Job type	Condition	*M* (*SD*)
Typical for women	Binary	2.13 (1.17)
	GII	2.27 (1.15)
Neutral	Binary	4.12 (1.30)
	GII	4.14 (1.19)
Typical for men	Binary	5.77 (1.22)
	GII	5.75 (1.07)

Predicting Interest in a Job

Though we had not found an effect of manipulation on interest ratings in Study 2, we wanted to explore whether this was to do with the design of Study 2. As such, in Study 3 (this study) we had changed our design to be between-subjects and include more potential moderators. Therefore, we investigated whether we could find an effect of our manipulation on participants' interest ratings of jobs, as well as checked whether order of questions, status rating of a job, gender typicality rating of a job, or gender essentialist beliefs, may moderate this effect. We expected different results for male and female participants, since gender-congruence is dependent on a participant's gender. To test these research questions, we ran two linear models (one for women and one for men; adjusted alpha level to account for multiple tests using Bonferroni correction; $p < .025$ is thus considered significant). We found no effect of manipulation on interest ratings, regardless of participant gender ($ps \geq .062$), replicating our null finding from Study 2. This suggests that interest in different jobs is not affected by GII manipulations.

Discussion

Our aims for Study 3 had been (1) to find differences in perceived societal beliefs and personal gender essentialist beliefs between people who had been subjected to a GII manipulation versus a binary manipulation, (2) to find differences in how much jobs are stereotyped by people subjected to a GII manipulation versus a binary manipulation, and (3) to explore potential differences in interest ratings of various jobs between people subjected to GII versus a binary manipulation. We found some evidence for our first two aims, but not for our third. Furthermore, we combined the materials of Studies 1 and 2 and were therefore able to compare them. We describe each of our findings, in turn, below.

For our first aim, we found that participants agreed with most of the statements about each video or article regarding perceived societal beliefs, with the exception of the GII article. Specifically, participants in the binary condition generally agreed that most people in society believed in natural differences between the sexes and that the different gender roles for men and women are justified. On the other hand, participants in the GII condition agreed that most people in society believed that gender and sex are not strictly binary, and some people fall outside of the norm. However, they were not convinced that most people in society believed that there are no differences between the sexes or that it is generally accepted for men and women to act differently from the norm. These were two non-significant items, which were about the GII article. Nonetheless, given that participants were convinced by some of the statements made in the GII video and article, which was the manipulation that is non-normative in society, we think we succeeded in manipulating participants' beliefs about society.

Furthermore, we wanted to investigate whether we manipulated gender essentialist beliefs, with beliefs about society acting as a mediator. However, we found no effect of manipulation on gender essentialist beliefs, and no effect of perceived societal beliefs, our manipulated variable, on gender essentialist beliefs. This was contrary to expectations, which we had based on the results of Study 1, where we found a mediation effect of explicit binary beliefs on gender essentialist beliefs. This discrepancy in results could be to do with the differences in variables. While in Study 1 we asked participants what they personally believed about gender, in Study 3 we asked participants how much they thought society at large would believe certain statements about gender. In Study 3, we chose to ask about beliefs of society rather than individual beliefs because it allowed us to test whether our manipulation had affected the social desirability of beliefs about

gender, rather than personal beliefs. However, it meant that a limitation of this study was the lack of direct replication of the effect of own beliefs about gender on gender essentialist beliefs. We also cannot say how much our manipulation affected participants' personal beliefs about gender being binary or a spectrum in Study 3. Since our gender essentialism measure was also about personal beliefs, it makes sense that this would be related to the personal belief about gender, rather than others' beliefs.

For our second aim, we wanted to investigate whether our manipulation had an effect on the gender typicality ratings of jobs, which represent how much participants stereotyped jobs as being typical for women or typical for men. We also wanted to check which variables might moderate that effect: gender essentialism, interest rating or status rating. We found a main effect of manipulation on gender typicality ratings, but only for jobs that are typically done by women, and not neutral jobs or jobs typically done by men. Specifically, we found that participants subjected to the GII manipulation considered jobs that are typical for women, generally, to be more neutral than participants subjected to the binary manipulation. Gender essentialist beliefs were a moderator of this effect, such that low gender essentialism was related to less binary (more neutral, closer to scale mid-point) gender typicality ratings. These findings are interesting because, while women are increasingly entering jobs that were historically typically done by men, men are entering jobs that are typical for women at a lower rate (England, 2010; Torre, 2018). Our findings thus suggest that GII manipulations, in combination with decreasing people's gender essentialist beliefs, could be a first step in addressing this issue by getting people to rethink their stereotypes about historically feminine jobs. Interest and status ratings were also moderators for the effect of manipulation on gender typicality

ratings of jobs done by women, such that higher interest in a job, and higher perceived status of a job, were related to less binary (more neutral) ratings of jobs. We think this is because participants who were more interested in a job, or thought of it as higher status, wanted to stereotype it less, perhaps because seeing a job as desirable is linked to thinking of it as more than "just" a typically feminine job.

Moreover, we wanted to explore whether there were any differences in interest ratings of jobs, depending on manipulation. Though we had not found any evidence for this in Study 2, we wanted to see whether we also found null effects in a different design. We also wanted to explore more potential moderators, and the design of this study made it possible to include more of these than in Study 2. We did not find any effects of manipulation on interest ratings of jobs, suggesting that interest in a job was not affected. This may be because interest is more of a trait than a state (Steyer et al., 1999), or because our manipulation was more successful at affecting beliefs about others (e.g., how suitable a job is for men or women) than beliefs about participants themselves (e.g., how suitable a job is for myself).

Lastly, we were able to compare the materials, in terms of manipulations and job titles, of Studies 1 and 2. Specifically, we found that participants agreed more with the items regarding societal beliefs that were related to the GII video (taken from Study 2), than the GII article (from Study 1), whereas we found no such discrepancy between the binary video (from Study 2) and the binary article (from Study 1). This suggests that the GII manipulation used in Study 2 was more convincing than the GII manipulation used in Study 1. This could be due to the subject matter: the article was about a neuroscientific study showing that brains are not binary, while the video showed a scientist talking about evidence against the binary on various domains, such as hormones, culture, identity, etc. This means that the video might have

been more convincing due to the amount of evidence against the binary presented. It could also be because a video is more engaging than an article (Singhal & Rogers, 2012) and therefore more convincing when it comes to a message that goes against the norm. Since both the binary article's and the binary video's message was according to the norms of society (which is still set up in a binary way), we may not have found such a discrepancy for this condition. Furthermore, by including the jobs from Study 1 (Kindergarten teacher and IT trainer) in the job list of Study 3, we found that participants rated these two jobs as similar status, but different gender typicality. This gives us some context for the findings in Study 1, which showed that participants were less interested in the Kindergarten job in the GII manipulation. It suggests that this finding was not related to the status of the jobs, as we had expected, and may instead have been related to the gender typicality of the jobs.

In conclusion, we were able to find some evidence for our first two research questions, regarding the effect of the manipulations on beliefs about society and on gender typicality ratings of jobs. We were not able to find evidence for the effect of the manipulation on interest ratings of jobs. We were able to compare materials from Studies 1 and 2 and therefore get more context about the results of each study.

General Discussion

Across three studies, our aims were to investigate the potential impact of GII, as compared to binary messaging, on gender binary beliefs, gender essentialist beliefs and stereotypes about jobs (Studies 1 and 3), as well as interest in jobs typically dominated by one gender group (all studies). We found some first evidence for the impact of GII on each of our measures except job interest, which may not be easily manipulated by GII. There were, however, some discrepancies in both methods and results across studies, which we describe, and provide potential explanations for, below.

In Studies 1 and 3, we investigated whether exposure to GII, in the form of an article or video including scientific explanations for gender and sex not being binary, affect people's beliefs about gender being binary, as compared to a binary manipulation. In Study 1, we measured whether participants thought gender was binary or a spectrum, thereby measuring their own beliefs. In Study 3, on the other hand, we asked participants to report how much they thought *other* members of society would agree that gender was (or was not) binary, thereby addressing the potential social desirability of a binary belief. In both studies, we found that our manipulation affected these gender binary beliefs. This is in line with theory, which states that gender binary beliefs are not innate or unchangeable and are instead culturally learnt (Hyde et al., 2019).

Furthermore, as Saguy et al. (2021) indicate, beliefs about gender being binary go hand in hand with beliefs that gender is a biologically essentialized category. As such, in Studies 1 and 3, we also measured whether exposure to GII would affect gender essentialist beliefs, as compared to a binary manipulation. In Study 1, we found no effect of manipulation on gender essentialist beliefs, instead finding that gender binary beliefs, which had been affected by our manipulation, further affected gender essentialist beliefs. This fits with the theory by Saguy et al. (2021), who propose a 'gender binary cycle' whereby binary gender ideology and biological-essentialist views of gender are in a feedback loop with each other. As such, a manipulation that disrupts binary gender beliefs would have an indirect (rather than direct) effect on gender essentialist beliefs, in the same way that a manipulation disrupting gender essentialist beliefs would have an indirect effect on gender binary beliefs. In Study 3, however, we did not find that gender binary beliefs further affected gender essentialist beliefs. We believe this is due to the measure for gender binary beliefs that we used in Study 3, which was about how much *others* in society

would agree with statements about gender being binary (or not), rather than about *personal* beliefs.

Similarly, disrupting gender binary ideology has been theorized to affect gender-based labelling and sorting (Saguy et al., 2021). In our case, we investigated the potential effects of exposure to GII, as compared to a binary manipulation, on how participants sort jobs as typical for women or typical for men. We did this in Study 1 using a measure of implicit associations of jobs with binary gender groups. In Study 3, we instead used an explicit measure, asking people to indicate how typical a job is for men or women. We found no evidence of our GII affecting participants implicit associations about jobs in Study 1. However, we did find evidence of our GII affecting participants explicit gender-sorting of jobs that are typical for women (but not neutral jobs or jobs typical for men). We think this discrepancy in results may be due to the asymmetry in how malleable implicit and explicit beliefs are. For example, a series of studies by Gregg et al. (2006) showed that implicit beliefs were easily induced but very hard to reverse, whereas explicit beliefs were more prone to being undone. Similarly, implicit stereotypes about jobs are induced from a young age (e.g., Canessa-Pollard et al., 2022; Garrett et al., 1977) and may therefore be too deep-seated to change with a single manipulation. The finding, however, that explicit beliefs about jobs typical for women are potentially malleable after exposure to a GII, is promising. While women are increasingly entering male-dominated jobs, men's participation in female-dominated fields is still low (England, 2010; Torre, 2018). As such, initiatives that have the potential to change perceptions of female-dominated jobs may help increase men's participation in such fields. However, given that we only investigated this effect of GII on explicit perceptions of jobs typical for women in one of our studies, replication is sorely needed to be able to confirm this finding.

Lastly, across all three studies, we investigated whether, along with stereotypes about jobs, GII have the potential to affect people's interest in jobs typically dominated by one gender group. Though in Study 1 we found some preliminary evidence for our GII manipulation affecting interest in a HEED job, as compared to a STEM job, in subsequent follow-up studies we found no effect of manipulation on interest in jobs of varying status and gender-typicality. We have two potential explanations for this null effect. First, interest in a job may be more of a trait than a state (Steyer et al., 1999), meaning it is less malleable or context-dependent. Second, there may be a self-other bias at play (Pronin et al., 2004). Interest in a job, as compared to beliefs about stereotypicality of jobs, is more about the self ('how suitable this job is for me') than other people ('how suitable is this job for men/women in general'). Self-perception may be less malleable than perception of others. This is also in line with our finding that gender identification (i.e., perception of one's own gender) consistently did not differ between people subjected to a GII manipulation and people subjected to a binary manipulation. On the other hand, participants' beliefs about other people's genders (i.e., whether gender is binary or not, in general) were affected by the manipulations. As such, a single GII manipulation may not be suitable for changing people's perceptions about their own interests (or their own gender identity). However, it is possible that consistent messaging that gender is not binary may, over time, affect people's interest in participating in jobs dominated by the 'other' gender. More research, perhaps in the form of a longitudinal study, is needed to investigate this.

Limitations

There were several limitations to the studies described in this paper. First, we used different designs and measures to investigate the same effect across studies, which constitutes a *conceptual* replication, and is a useful tool for investigating whether a certain result

is dependent on a particular methodology. However, our lack of *direct* replication across studies is a limitation. This is because it makes it impossible to say whether a discrepancy in results is due to differing methodology or due to statistical error. Considering we found some discrepancies in results, we recommend that future researchers use direct replication to be able to further contextualise the results described in this paper. Moreover, in this paper we have researched a single type of GII, namely media segments (articles or videos) describing reasons why gender is not binary. This represents only one type of GII that is on the rise throughout the world, and other types of initiatives (such as gender-inclusive language or toilets) may affect people's beliefs differently. In fact, some initiatives may be particularly threatening to people, because they disrupt the gender system (Morgenroth & Ryan, 2021). Given that social identity threat may be related to resistance to change (Scheepers et al., 2009), initiatives that people perceive to be more threatening may affect people's beliefs to a lesser extent. Since we did not include a measure for threat in our studies, we cannot say how threatening participants found our manipulations and therefore how this might have related to their resistance to change their beliefs. As such, we recommend that future investigation into how GII may affect people's beliefs include more types of GII to compare to each other and include a measure for psychological threat to such initiatives.

Practical Implications

There are a few practical implications to the findings described in this paper. First, our findings that GII affect gender binary beliefs indicate that GII may indeed be a useful tool for increasing inclusion of gender minorities (whose existence is still often denied; e.g., Nadal et al., 2012). Second, our findings that GII may indirectly affect gender essentialist beliefs suggest that GII may help reduce gender stereotypes by addressing beliefs that men and women are fundamentally different.

Third, we found preliminary evidence that GII have the potential to decrease stereotypes about female-dominated jobs and may therefore represent another step towards decreasing the gender labour gap (OECD, 2021). We therefore think that, besides the important benefit of increasing the inclusion of transgender and non-binary individuals, the implementation of GII can also benefit cisgender women and men.

Conclusion

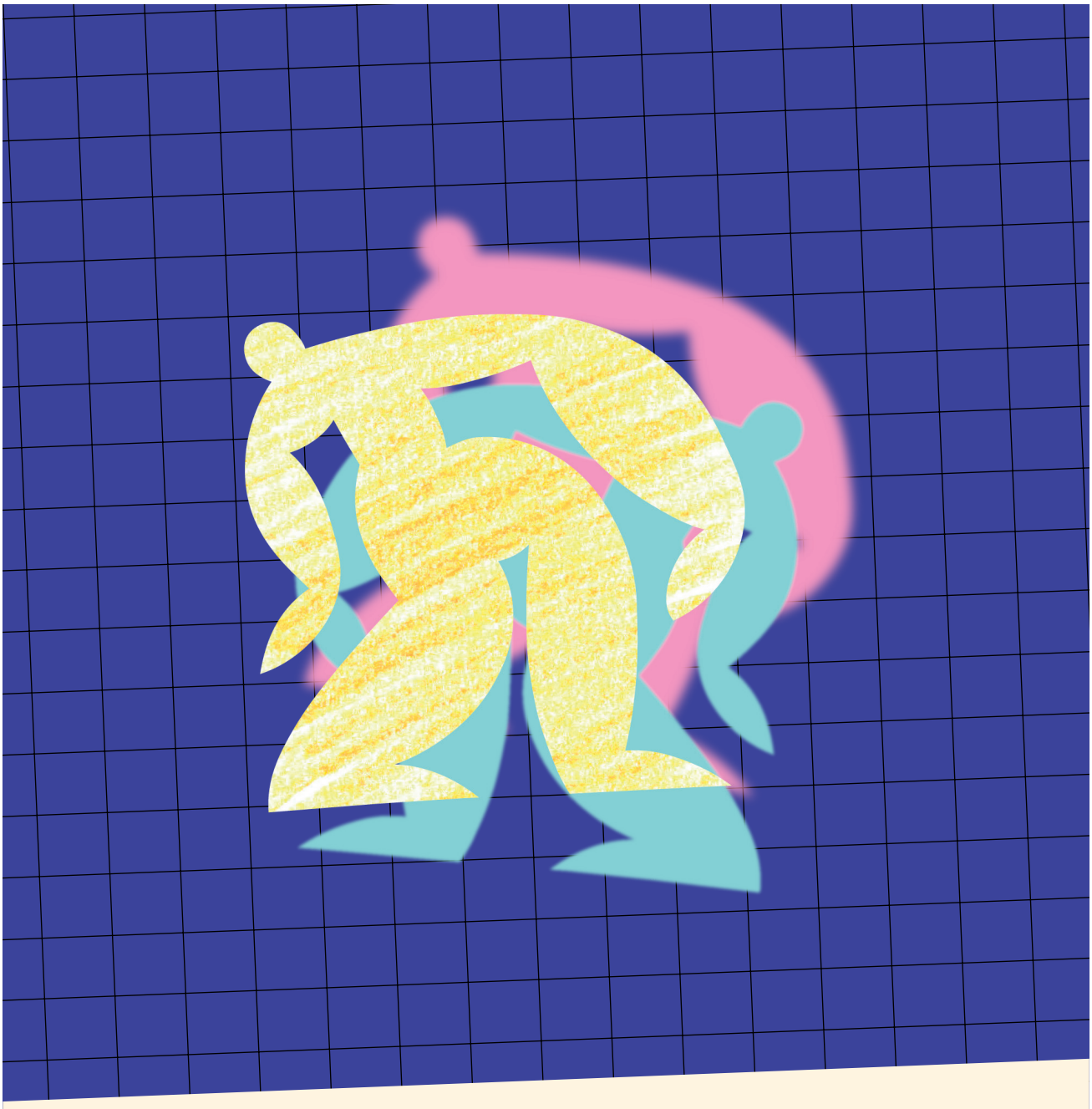
The studies described in this paper are a step towards investigating the potential positive effects of *gender-inclusive initiatives* (GII), which raise awareness about gender being a spectrum, rather than binary. We show that these positive effects may not only be for non-binary and transgender people, but there may also be benefits of implementing GII for cisgender women and men. Specifically, GII may disrupt the gender-binary cycle and increase the belief in non-binary genders, as well as decrease beliefs in gender stereotypes.

Footnotes

[1] For information on why we chose an IRAP, as opposed to an Implicit Assessment Test (IAT), please see Supplementary Materials.

[2] In a series of one-sample t-tests with Bonferroni correction, all jobs except Fashion consultant and Administrative secretary were found to be significantly different from scale mid-point (true medium status), meaning all jobs were seen as either high or low status to some extent. However, for replication purposes, we classified more jobs as medium status to be able to compare results from Studies 2 and 3.

[3] In a series of one-sample t-tests with Bonferroni correction, all jobs were found to be significantly different from scale mid-point (true neutral), meaning all jobs were seen as somewhat gendered. However, for replication purposes, we classified some jobs as neutral to be able to compare results from Studies 2 and 3.



Part 2:
Gender (Self-)Categorisation

Chapter 4: Gender Identification Beyond the Binary and its Consequences for Social Well-Being

"Girls will be boys and boys will be girls. It's a mixed up, muddled up, shook up world"

Lola - The Kinks

Author contributions: Miriam I. Wickham (Conceptualisation; Data curation; Formal analysis; Investigation; Methodology; Validation; Writing – original draft; Writing – review & editing); Félice van Nunspeet (Conceptualisation; Methodology; Validation; Writing – review & editing); Naomi Ellemers (Conceptualisation; Funding acquisition; Methodology; Validation; Writing – review & editing)

This chapter is published under: Wickham, M. I., van Nunspeet, F., & Ellemers, N. (2023). Gender Identification Beyond the Binary and Its Consequences for Social Well-Being. *Archives of Sexual Behavior, 52*(3), 1073-1093. https://doi.org/10.1007/s10508-022-02453-x

Abstract

Recent societal initiatives (e.g., gender-neutral toilets, clothing, and language) highlight the ongoing shift of gender away from binary categories: "man" and "woman." We identified and investigated two reasons for this shift: that many people may not identify with strictly binary categories, and that this may have negative social consequences. Employing a multiple-identification model, we measured intergroup self-categorisation with both men and women (Studies 1 and 2), as well as with a "third gender" (Study 3), and investigated how multiple identifications are related to social well-being (Studies 2 and 3). In Study 1 (N = 182, mean age = 32.74, 121 women), we found that a binary model was not the best fit for our gender identification data. In Study 2 (N = 482, mean age = 30.98, 240 AFABs), we found four clusters of gender identification, replicating previous research. Furthermore, we found that gender non-conforming participants reported being less able to be their authentic selves than binary participants. We also found that participants who identified lowly with both binary genders reported lower well-being in general (belongingness, self-esteem, life satisfaction, positive affect). In Study 3 (N = 280, mean age = 36.97, 140 AFABs), we found that asking about a third gender seemed to change how much participants reported identifying with men and women. We also found that gender non-conforming participants reported lower authenticity, belongingness, and self-esteem. We conclude that moving away from binary categories of gender may be beneficial to many non-conforming people of different nationalities, including cisgender, heterosexual people.

Keywords: *gender identification, gender non-conformity, non-binary, social well-being, authenticity*

Throughout the developed world, there has been an increase in legal recognition of a "third" gender/sex category, which is separate from male or female (e.g., in Australia: Bennett, 2014; in Germany: Dunne & Mulder, 2018). The legal recognition of more than two genders (e.g., "man," "woman" and "non-binary") is an example of the current shift away from the "gender binary," where there are only two gender categories, and toward a "gender spectrum," to acknowledge that there are multiple ways for people to define their gender. Recent estimates suggest that anywhere from 3% to 35% of the general population see themselves as part man and part woman, or as something completely different from man and woman (in the Netherlands: Kuyper & Wijsen, 2014; in Flanders, Belgium: Van Caenegem et al., 2015; in Israel: Joel et al., 2014). We set out to further examine these different options to indicate gender, by allowing more freedom for expression of gender identification (GI) in our questionnaire items. We suspected that this would reveal a larger percentage of gender non-conformity in the general population than some of the previous work. Furthermore, we argue that being gender non-conforming in a society with a binary narrative would have negative consequences for one's social well-being. Across three studies we therefore tested the implications of gender non-conformity on social well-being, specifically, societal inclusion, general self-esteem, and life satisfaction.

Toward a More Inclusive Gender Conceptualization

Our world is highly binary in terms of gender. The gender binary is learnt from a young age, for instance, when teachers implicitly treat boys and girls differently (e.g., Duffy et al., 2001; Tsouroufli, 2002), and binary gender labels on toys affect children's play behaviour (Yeung & Wong, 2018). When asked to indicate gender in registration of person characteristics, most forms provide only two options (Morgenroth & Ryan, 2018) and the binary gender categories are (sometimes falsely) considered biologically essential and meaningful (Skewes et al., 2018).

In social psychology, research into GI is a little more nuanced, in that participants are generally asked to state the *extent* to which they identify with their binary gender group, based on which we classify low or high identifiers (e.g., Derks et al., 2011). However, such an approach is still binary because a participant of the female sex is asked only to indicate how much they identify with women, thereby disregarding other-sex GI. This is an example of how most of the psychological literature regards gender identification as a binary construct, though not all, as will be explained.

Importantly, a large body of research has investigated psychological androgyny in terms of personality traits. This research began with Bem (1974) and has impacted gender research greatly, by promoting equality between men and women, showing the malleability of gender roles, and underlining the negative effects of polarization of gender groups (Dean & Tate, 2017). This body of work was an important steppingstone for research into gender beyond the restrictive, binary conceptualization that is still prevalent today. The foci of this research were gender *roles*, as measured through personality traits. For instance, women are typically stereotyped to be caring and tender, while men are typically thought of as agentic and dominant, and these stereotypes have persisted across decades (Bem, 1974; Haines & Stroessner, 2019). The present paper was inspired by this important literature, but our focus is on gender *identification*, meaning self-categorisation as a member of different social (gender) groups (Turner & Reynolds, 2011). While gender *roles* are about the way that people are stereotyped in terms of personality traits, gender *identification* is about feelings of group membership. We also show that these two concepts are distinct in our own data; the results of this are beyond the scope of this paper but can be found in the Supporting Information.

The idea that gender is a binary construct is currently changing. We can see this in the legal recognition of non-binary genders (e.g., in Australia: Bennett, 2014; in Germany: Dunne & Mulder, 2018), and in the increase of gender-inclusive interventions, such as gender-neutral toilets (Gershenson, 2010), unisex clothing lines (Park, 2014), and gender-neutral language (Hord, 2016), among others. In this paper, we identified two possible reasons for this change. First, people may feel that the gender binary does not reflect the way they see their own gender. Second, the gender binary restricts people in such a way that their well-being is compromised. As such, we tested gender identification (in terms of intergroup self-categorisation) in the general population, and how this relates to social well-being. We describe each of these research questions below.

Gender Identification Beyond the Binary

While one might assume that most cisgender, heterosexual (i.e., normative) individuals have much less complex gender identities than people belonging to the LGBTQIA+ conglomerate, research suggests that this is not the case. This means that the gender binary may not be a good reflection of the way people in the general population identify. Cisgender, heterosexual participants in an interview study showed much diversity in how they perceived their own gender and sex (Abed et al., 2019). Various studies have employed a multiple-identification model of gender and found that both normative and non-normative individuals can identify with various genders at once. A multiple-identification model is one in which participants, regardless of their explicit social category (e.g., "woman"), are asked how much they identify with several categories (e.g., "man" and "woman") of the same construct (e.g., gender). For instance, Joel et al. (2014) asked Israeli participants how much they felt like a woman, a man, both or neither in the past twelve months. They found that some (self-assigned) men felt more

like a woman than a man, and vice versa, and that approximately 35% of men and women felt like both a man and a woman. Similarly, Martin et al. (2017) asked US children (ages 5-10 years; whose gender identity and sexuality is not fully formed yet) how similar they felt to members of their own assigned gender at birth (AGAB) as well as how similar they felt to members of the "other" gender. They found that children could be clustered into four distinct categories: Own-Gender Similar (high identification only with own AGAB), Cross-Gender Similar (high identification only with "other" gender), Both-Gender Similar (high identification with both binary genders) and Low-Gender Similar (low identification with both binary genders). They estimated that 30% of children were Both-Gender Similar. Various studies have used Martin and colleagues' (2017) gender similarity measure and largely replicated their results in Dutch, Italian and US samples of young adults (Andrews et al., 2019; Baiocco et al., 2022; Endendijk et al., 2019). Similarly, Pauletti et al. (2017) found relatively low correlations between own-gender typicality and other-gender typicality in children, further showing that these are two distinct concepts.

Similarly, using a multiple-identification model, there is a small body of literature examining so-called "gender ambivalence" within the Dutch and Flemish context. Gender ambivalence is a term coined by Kuyper and Wijsen (2014) to describe identification that is approximately equal for both binary gender groups. A gender ambivalent person therefore identifies with women to a similar extent as they identify with men. Kuyper and Wijsen (2014), and Van Caenegem et al. (2015) estimated that in the Netherlands and Belgium, respectively, 3-5% of people are gender ambivalent. Gender ambivalence is a form of "gender non-conformity," the term we use throughout this paper.

In the present paper, we tackled the research question of how both cisgender and heterosexual, as well as LGBTQIA+ adults identify in

terms of their gender, by combining methods from these four previous studies, and adding to them. Previous papers differed from each other in various aspects, and in this section, we address each of these aspects, and describe how we tackled them in our own studies.

While most of the previous studies each included one measure of social gender identification, consisting of one item (in Joel et al., 2014, "feeling like a man/woman;" in Kuyper & Wijsen, 2014, and Van Caenegem et al., 2015, "experiencing oneself as a man/woman"), Martin et al. (2017) and Pauletti et al. (2017) measured gender identification using scales including several items. We argue that to measure as complex a concept as gender identification, one must allow the participant to reflect on various aspects of their gender, by including several items that capture the diverse ways in which participants may feel about their gender.

Moreover, in one study identification with "women," "men," "both," and "neither" was each measured using separate items (Joel et al., 2014). However, most other studies measured identification with "women," and with "men," and inferred which participants identified with both or neither from the data (e.g., those participants scoring low on identification with women and low on identification with men, were considered to identify with neither; Andrews et al., 2019; Baiocco et al., 2022; Endendijk et al., 2019; Kuyper & Wijsen, 2014; Martin et al., 2017; Pauletti et al., 2017; Van Caenegem et al., 2015). We argue that either method is valid for assessing identification with "both" and "neither" men and/nor women, however, we favour simplifying the participant's experience in answering questions where possible. Consistent with work by Kuyper and Wijsen (2014), Martin et al. (2017), and Van Caenegem et al. (2015), we thus measure identification with "men" and with "women" (and a third gender in Study 3), and do not additionally ask participants to indicate how much they identify with "both" or "neither."

Furthermore, in previous studies, identification was measured in relation to binary groups ("men" and "women"), and genders outside of the binary were not included in the model. Such a model is less binary than other work, where participants are asked about only one binary gender group. However, since all questions relate back to the binary gender groups it implies that one cannot identify with a gender that is unrelated to men and women. We argue that it is important to explore how participants' responses may differ if asked only about binary genders, or about binary and non-binary genders.

Each of these former studies also used different methods, in line with their measures, for calculating estimates of the prevalence of gender non-conformity, or the amount of people who identify with multiple gender identities (e.g., in Martin et al., 2017, a cluster analysis; in Joel et al., 2014, frequencies divided by gender group; in Kuyper & Wijsen, 2014, and Van Caenegem et al., 2015, descriptive statistics across the whole sample). Given that we argue for more complex questionnaires with diversity in items for measuring identification, we favour a cluster analysis, which allows us to display the complexity of the data in a simplified manner for the reader. This methodology may seem reminiscent of Bem's (1974) approach to measuring and scoring psychological androgyny. This approach displayed masculinity and femininity as orthogonal dimensions rather than a linear construct with two endpoints and categorised people into groups (e.g., psychologically androgynous) according to their femininity and masculinity scores (or rather, the difference score between them). We similarly argue that gender identification with various gender groups should be seen as orthogonal and not a linear construct with two endpoints. Using cluster analyses, gender identification scores with various gender groups can be seen as statistically independent. We clustered participants into groups based on those scores for the

purposes of giving estimates and conducting further analyses only, and not to claim that the clusters we find should be seen as the new norm for categorising gender. However, we also argue that cross-validation of measures is important, to show that similar results can be obtained, regardless of the method used.

Lastly, in each study the participants shared a country of residence: the USA (Andrews et al., 2019; Martin et al., 2017; Pauletti et al., 2017), Israel (Joel et al., 2014), Italy (Baiocco et al., 2022) the Netherlands (Endendijk et al., 2019; Kuyper & Wijsen, 2014) and Belgium (Van Caenegem et al., 2015). The redefinition of gender as a non-binary construct is happening world-wide, and to compare whether participants in different studies experience their genders in similar ways, a multinational sample is necessary.

Reduced Well-Being

Importantly, previous work does not draw a link between gender non-conformity and well-being. Social identification, including gender, affects social well-being because some social categories are more discriminated against than others. Since the societal norm is that "gender is binary" and this is perpetuated from a young age, it is likely that people consider gender non-conformity to be an undesirable or uncommon trait, which would have negative consequences for gender non-conforming people. For example, it has been found that people feel negatively about physically androgynous people, who are difficult to categorise in a binary way (e.g., Stern & Rule, 2018). A highly binary society also means little representation of gender non-conformity in politics or media, which can cause negative feelings in gender non-conforming people (Klasen, 2007). Gender non-conforming people may feel invisibly dissimilar from other people, whom they assume to be binary, which can reduce feelings of inclusion (Şahin et al., 2019). Because of the perceived deviance from the social norm, we expected

that gender non-conforming individuals would report lower social well-being. Such a finding would underline the need to move away from binary conceptualizations of gender in society.

We were interested in social well-being, rather than clinical well-being. Examinations of clinical symptoms, such as gender dysphoria, depression, and anxiety, in people who identify with a non-normative gender, can be found in earlier work (e.g., Dhejne et al., 2016; Kuyper & Wijsen, 2014). This is important work, as it shows that there are serious clinical implications to non-normative gender identity. However, one does not have to have a clinical diagnosis to be suffering from the consequences of being non-normative in a binary society. People with non-normative gender identifications may feel less included by society, because they feel dissimilar from the social norm (Kristof-Brown et al., 2005). Specifically, they may feel that they are not allowed to be their authentic (non-normative) selves, or that they do not belong with other (binary) people (Jansen et al., 2014). Because their personal identification does not fit with the collective, people may have lower personal self-esteem, meaning lower acceptance of one's own personal identification (Rosenberg, 1979). They may also feel that their general quality of life is impacted by their lowered inclusion, meaning they are less satisfied with their lives (Diener et al., 1985). Lastly, their emotional profile might be impacted negatively, without necessarily fitting the diagnostic criteria for a mood disorder, such that they feel more negative affect and less positive affect in their lives (Andrews & Withey, 1976). All of these measures are related to one another and are part of well-being as a member of a social environment. In our studies, we measured social well-being in terms of feelings of inclusion in society as well as general self-esteem, life satisfaction, and positive and negative affect, and hypothesized that we would find differences between gender conforming and gender non-conforming participants in all of these measures.

Social well-being, while being affected by one's (private, internal) gender identification, is also affected by how one is categorised by others (externally). One may identify as both man and woman but be perceived solely as a man. This is related to biological features that are perceived as male or female, which tend to correlate with one's assigned sex at birth (ASAB). It can also be related to how one expresses themselves, for instance one's name, clothes, or pronouns, which tends to correlate with one's self-assigned gender identity. Men tend to have more status in society than women (e.g., Eagly, 1983), and therefore feel more included by society in general. As such, being perceived as a man may increase feelings of social well-being. Self-assigned gender identity and ASAB are, therefore, potential moderators for the relationship between gender (ambivalent) identification and social well-being, which we tested.

The Present Studies

In our studies, we estimated the prevalence of gender non-conformity across three multinational samples by employing a multiple-identification model, and using a questionnaire that measures self-categorisation as a member of a social group. We hypothesized that we would find four clusters of GI, in line with Martin et al. (2017). We hypothesized that at least one of those clusters could be considered gender non-conforming (in Studies 1 and 2, identifying the same amount with both gender groups; in Study 3, identifying to a certain extent with a third gender). By using this methodology, we expected to find a higher percentage of gender non-conforming people than previous estimates reported by Kuyper and Wijsen (2014) and Van Caenegem et al. (2015).

We measured intergroup self-categorisation (Turner & Reynolds, 2011), which is the cognitive component of social identification (Ellemers et al., 1999), using a scale that encompasses various aspects of self-categorisation: identification as a member of the group, collective

self-esteem (i.e., how important a group membership is to the self), and similarity to the group. We argue that using such a scale provides the participant with freedom to reflect on various parts of their gender self-categorisation. We also argue that, while Self-Categorisation Theory (Turner & Reynolds, 2011) poses that individuals define themselves as members of *distinct* social groups, in the case of gender, the binary groups of "men" and "women" may not be entirely distinct, having somewhat blurry social boundaries. By asking people how much they self-categorise as men and women, we aim to show that participants acknowledge these blurry social boundaries and can identify with both, with neither, or with some other gender entirely. Additionally, to cross-validate our gender self-categorisation scale, which relies on participants' understanding of language, we compared the results of the scale with the results of an alternative, pictorial (i.e., less reliant on language) measure of self-categorisation, which we included in Study 3.

To test our prediction that gender non-conformity, as compared to binary identification, would be related to lower social well-being, in Studies 2 and 3 we included measures for societal inclusion, general life-satisfaction and self-esteem. In Study 2 we additionally measured positive and negative affect when reflecting on those measures. We also tested the potential moderation effect of ASAB and self-assigned gender identity on the relationship between gender non-conformity and social well-being.

Study 1

Method

Participants

Based on a medium effect size (approx. Cohen's $f = 0.25$), as is typical in social psychology in general (Richard et al., 2003), and the expected number of GI categories being four, we calculated that our *N* would need to equal at least 179 participants to achieve 80% power[1].

We recruited 182 participants, whose age ranged from 18 to 64 (M = 32.74, SD = 10.05). Sixty participants said that their gender was male; 121 participants said they were female, and one person indicated being "Other." Our sample was predominantly living in the United Kingdom (55.9%), as is usual on Prolific, with the rest of participants spread out across eighteen other (mostly Westernized) countries world-wide. For (more) information about participants' country of residence, education level, employment situations, level of feminist identification or LGBT+ identification, please see the Supporting Information. It should be noted that the results of this study are not affected by any of these demographic profiles.[2]

Materials

Gender Identification. To test the extent of self-categorisation as each gender group, seven items were administered about each gender group.[3] Items included "I identify with (other) women/men" and "I feel that I belong to the group of women/men" (for all items, please see Supporting Information). All items were administered twice, with the items being identical in both cases, other than the replacement of the word "men" with "women" and vice versa (items about men α = .86, items about women α = .91; for factor analysis, see Supporting Information). They were answered on a Likert scale going from 1 (highly disagree) to 7 (highly agree). For gender identification with women, we took the mean of the items about women, while for gender identification with men, we took the mean of the items about men. We used mean scores, rather than individual items, in order to be able to treat these as continuous variables, and to be able to measure more variance in gender identification.

Procedure

Participants were recruited using the online platform Prolific. They were redirected from Prolific to the survey's Qualtrics link and

asked to answer all questions. Participants were first asked about their gender identification, with items shown in a random order. They were additionally asked to answer questions of the following scales, which are beyond the scope of the current paper[4]: Bem Sex Role Inventory, two questionnaires about typically feminine and typically masculine norms, belief in "gender-as-binary" or "gender-as-spectrum" questions. Lastly, they were asked demographic questions. They were shown a debriefing and given the opportunity to retract their data. Participants received payment for their participation.

Results

Gender Identification

We created two variables: gender identification with women (GIW) and gender identification with men (GIM). We found a significant difference in GIM and GIW; respectively M = 3.64, S.D. = 1.28, min. = 1.00, max. = 7.00; M = 4.29, S.D. = 1.37, min. = 1.00, max. = 7.00; $t(181) = 4.24$, $p < .01$, Cohen's d = 0.29; indicating that our sample on the whole had a higher GIW than GIM. This may be due to the higher number of female participants in the sample.

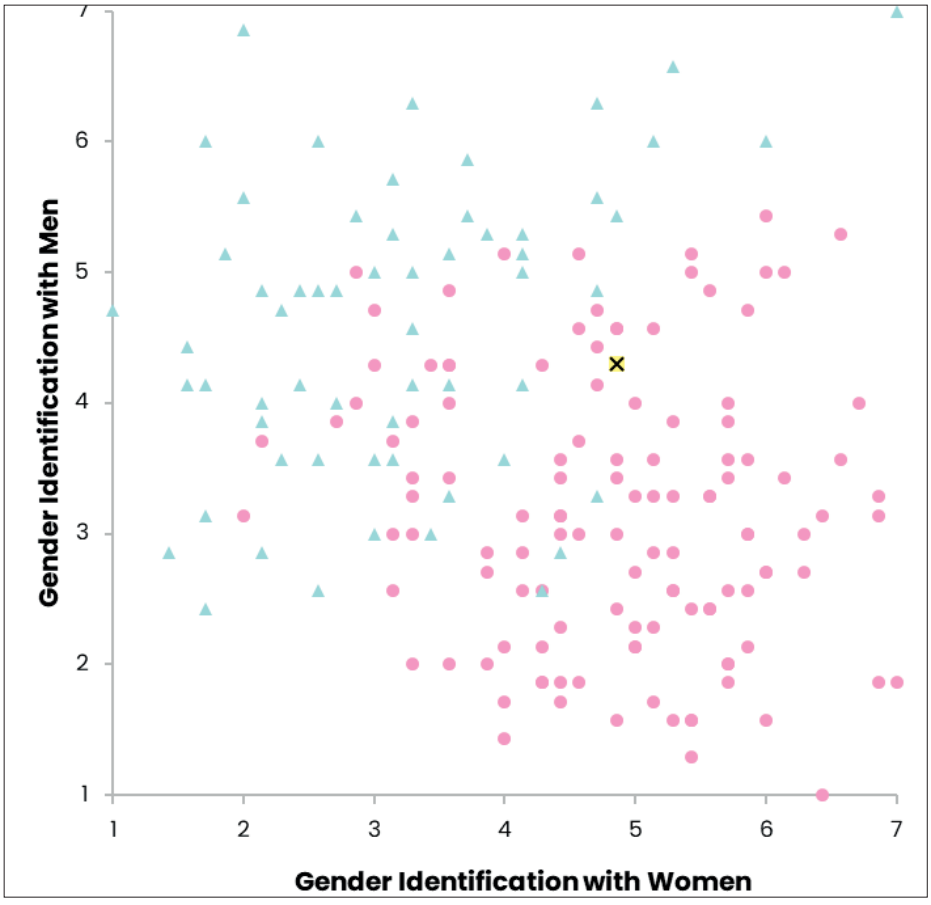
Relationship Between Gender Identification with Women and Gender Identification with Men

While we established that our GIW items loaded onto a different factor than our GIM items, thereby suggesting that they are two different concepts, we wanted to test how much they correlated with each other. While two concepts can be distinguishable from one another, they may still be highly related. In the case where GI is a spectrum with two opposite ends (male and female), one would expect that GIM would be moderately to highly negatively correlated with GIW. To test this we ran a Pearson's correlation which revealed only a small negative correlation between GIM and GIW ($r(180) = -.26$, $p < .01$, see Figure 1). This indicates

that, while GIW is negatively correlated with GIM, they cannot be considered complete opposite ends of the same spectrum.[5]

Figure 1

Scatter Plot of Gender Identification Results from Study 1



Note. Triangles represent self-reported men, circles represent self-reported women and the X represents the self-reported Other participant.

Given the low correlation between GIM and GIW, we wanted to test how many clusters of GI participants would fall into. To test this, two-step cluster analyses with two inputs (GIM and GIW) and different numbers of clusters (two, three, four and five) were run. In terms of silhouette of cohesion and separation, the qualities of the 2-cluster, the 3-cluster and the 4-cluster model, but not the 5-cluster model, were all similar and good (silhouette ≥ .5) Most importantly, Bayesian Information Criteria of the different cluster models (BICs, used to compare different models with each other), showed that splitting the participants up into three clusters resulted in better (lower) BICs than the 2-, 4- or 5-cluster models; 3-cluster model *BIC* = 174.68, *BIC* change = 23.03. Thus, a binary model of gender is not the best fit, and participants could best be categorised into three GI categories. This is one less than hypothesized based on previous literature (Martin et al., 2017). Based on the cluster centres (see Table 1), we suggest that the missing cluster is a low male, low female identifying cluster.

Table 1

Cluster Centres for GIW and GIM

	Cluster 1 (high male, low female identifiers)	Cluster 2 (high male, high female identifiers)	Cluster 3 (high female, low male identifiers)
Cluster center for GIW	2.74	5.04	5.13
Cluster center for GIM	4.18	4.95	2.56

Note. Absolute range of GIM and GIW: 1-7.

Discussion

Our goals for the first study were a) to test whether a binary or dual-identification model of gender fit the adult people's identifications

better, and b) to add to the literature about gender non-conformity by providing another estimate of the occurrence of gender non-conformity, using our methods.

We found that a dual-identification model of gender fits our data better than a binary model of gender. Specifically, we found three clusters of GI, one less than previously found by Martin et al. (2017) in children. Martin et al. (2017) had found that the four gender clusters were male identifiers, female identifiers, identifiers of both, and identifiers of neither (low male as well as low female identification). It seems that, in our data, we did not find low identifiers of both genders. This may be due to the development of GI between childhood and adulthood, but we explore this further in Study 2.

A limitation to this study was the overrepresentation of women over men, which caused us to refrain from providing an estimate of the occurrence of gender non-conformity as this would likely be skewed. To circumvent that in Study 2, we recruited equal numbers of people assigned male and female at birth. As such, a main aim of Study 2 was to report percentages of people in each cluster of the model, given a sex-equal sample.

Furthermore, our sample in Study 1 was mostly living in the UK or in the Westernized world, while changes in the gender narrative are happening world-wide. We therefore wanted to recruit a more international sample in Study 2 in terms of country of residence.

Lastly, we were interested in the relationship between GI and social well-being. In Study 1, we had not included any measure of social well-being. As such, in Study 2, we measured well-being, more specifically, in terms of societal inclusion (authenticity and belonging), general self-esteem, life satisfaction and affect (positive and negative). Our hypothesis was that gender non-conformity would be related to

lower feelings of well-being because the normative system is the gender binary, and gender non-conforming people may not feel that they "fit" into the normative system. We also explored some potential mediators and moderators for the relationship between GI and well-being, namely ASAB, and self-assigned gender identity.

Study 2

Method

Participants

Based on the small effect sizes found in Study 1 and based on previous literature (Goldman & Kernis, 2002; Richard et al., 2003) we expected our effect sizes to be medium-small (approx. $f = 0.15$). Based on Study 1, we expected to find three GI categories, however, because of the underrepresentation of men in that sample, we kept the possibility of finding four groups (like Martin et al., 2017) open. We would run ANOVAs to measure differences in six measures of well-being. Given this, we calculated a required N of 450 participants to achieve 80% power.[6] We recruited 482 participants with an equal number of AMABs and AFABs ($N = 241$ each), using Prolific's pre-screening feature (participants were asked the following question: **"What sex were you assigned at birth, such as on an original birth certificate?"**). Two-hundred and forty-two participants identified as men, 240 identified as women, and none identified as other. Participants' age ranged from 18 to 76 ($M = 30.98$, $SD = 9.94$). Additionally, since our sample in Study 1 was predominantly White, while the inclusion of non-binary genders is a phenomenon that crosses the boundaries of ethnic privilege, we also took care to recruit an equal number of White and non-White participants ($N = 241$ each). This was also done using Prolific's pre-screening (participants were asked the following question: **"What is your ethnicity?"**[7]). Our sample in Study 2 was also more diverse in

terms of country of residence than in Study 1. Participants lived in 36 different countries world-wide (not all of which Westernized), with a majority living in the UK (38.40%) and the US (19.71%). For more information about participants' ethnicity, education, employment, special needs, and country of residence, see Supporting Information. We also controlled for these demographics and, as in Study 1, did not find that any of them affected the results reported in this manuscript.

Materials

Gender Identification. We used the gender identification scale from Study 1 but had to exclude four items (two about men and two about women) due to high cross-factor loadings (see Supporting Information)[8]. We therefore measured GI using 10 items (five about men, $\alpha = .92$; five about women, $\alpha = .90$).

Social Well-Being.

Inclusion. We administered Jansen et al.'s (2014) inclusion scale developed to assess organizational inclusion and adapted the items to refer to "members of my society." There were 16 items; they included "Members of my society give me the feeling that I belong" and "Members of my society allow me to be authentic." Eight of the items were about belongingness ($\alpha = .97$); the other eight were about authenticity ($\alpha = .96$).

Self-Esteem. To measure general self-esteem, we administered 10 items which included "On the whole, I am satisfied with myself" and "At times I think I am no good at all" (reverse coded; Rosenberg, 1979; $\alpha = .92$).

Life Satisfaction. To measure general life satisfaction, we administered five items which included "In most ways my life is close to my ideal" and "If I could live my life over, I would change almost nothing" (Diener et al., 1985; $\alpha = .85$).

Positive And Negative Affect Schedule. This scale consists of 20 emotion words, 10 of which are positively valenced (e.g., "inspired") and 10 are negative (e.g., "ashamed;" Watson et al., 1988). We asked participants to reflect on how they felt while answering the questions about their gender, societal inclusion and well-being, and to rate how much they felt a certain emotion from the Positive and Negative Affect Schedule (PANAS) during that time (positive items $\alpha = .88$, negative items $\alpha = .93$).

Procedure

Participants were recruited through Prolific, using restrictions to recruit equal numbers of people in terms of ASAB and ethnicity. They were redirected from Prolific to the survey's Qualtrics link and asked to answer all questions. First they were asked to state their self-assigned gender. Next, they answered questions about GI, followed by the societal inclusion questionnaire, followed by the self-esteem, life satisfaction and emotion questionnaires. There were several extra measures that were administered to participants, the data of which we do not report in the paper as it is beyond the scope of this article: need to belong to ethnic and gender groups, beliefs about gender as a construct, and questions about a recent BBC article (see Supporting Information). The items within each questionnaire were administered in a random order. Last, participants answered demographic questions. They read a debriefing and had a chance to retract their data from the study. Participants received payment for their participation.

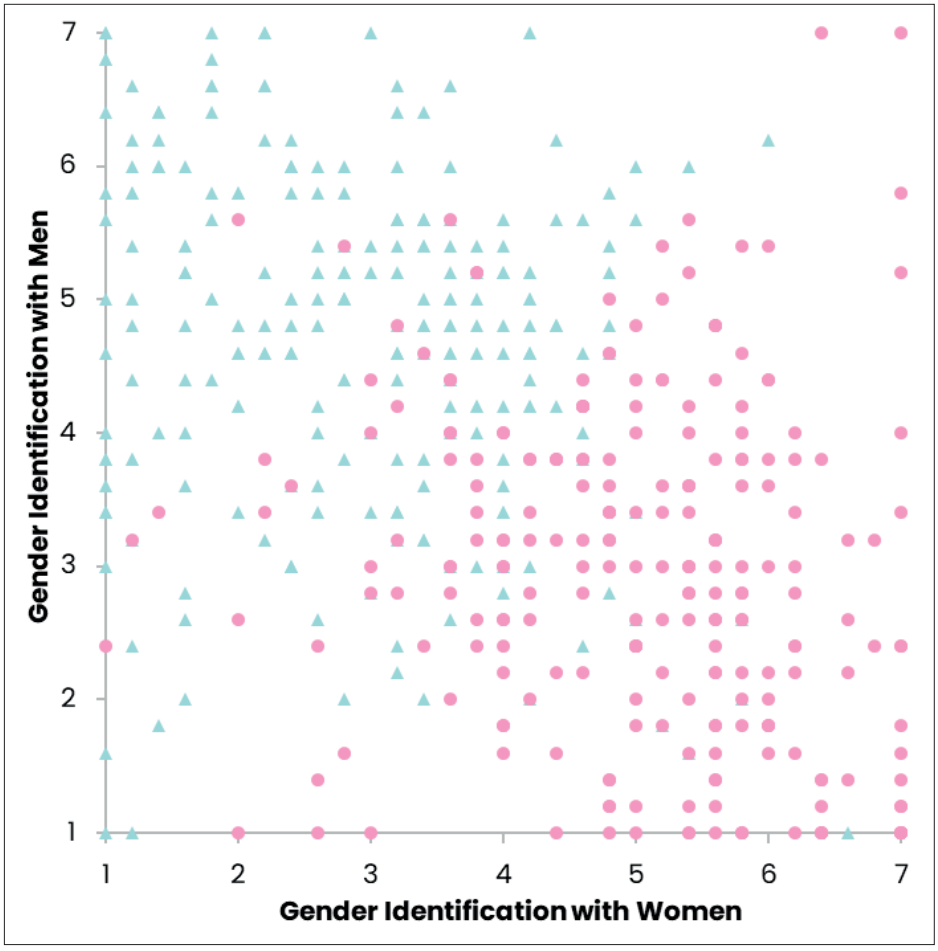
Results

Gender Identification

Mean scores of GIW and GIM did not significantly differ from one another; $t(481) = 1.27$, $p = .21$, M of GIW = 3.96, M of GIM = 3.80. A Pearson's correlation revealed that the correlation between GIM and GIW was small to moderate and negative ($r(480) = -.47$, $p < .01$, see Figure 2).

Figure 2

Scatter Plot of Gender Identification Results from Study 2



Note. Triangles represent participants assigned male at birth, circles represent participants assigned female at birth.

Several two-step cluster analysis with two inputs (GIW and GIM) revealed that a 4-cluster model (as compared to 2, 3, 5 and 6-cluster models) was the best fit in terms of percentage of variance, Bayesian Information Criterion (*BIC* = 340.54) and silhouette of cohesion (silhouette ≥ .5) and separation combined. This is different from the

3-cluster model found in Study 1. A four-cluster model was therefore chosen and the cluster membership of each participant was saved in order to perform further analyses. The clusters can be translated into: 1) male identifiers, 2) low identifiers (who identify lowly with both binary genders), 3) dual identifiers (who identify relatively highly with both binary genders), and 4) female identifiers. Table 2 shows the cluster centres as well as percentages of overall sample that were placed in each cluster. It seems that the cluster which was not found in Study 1 (low identifiers of both genders) could be found in the more sex-equal dataset of Study 2, in line with previous research (Martin et al., 2017).

Table 2

Cluster Centres for GIW and GIM, and Percentage of Participants Who Fell into each Cluster

	Cluster 1 (male identifiers)	Cluster 2 (low identifiers)	Cluster 3 (dual identifiers)	Cluster 4 (female identifiers)
Cluster centre for GIW	2.32	2.74	4.69	5.73
Cluster centre for GIM	5.52	2.93	4.42	1.99
Percentage of overall sample	25.31%	19.92%	30.08%	24.69%

Note. Absolute range of GIM and GIW: 1-7.

Given our multinational sample, and that gender is culturally constructed (e.g., Newman, 1995), we wanted to test whether participants with different countries of residence differed in their cluster memberships. A chi square test revealed no significant differences (X^2 = 117.82, p = .19), suggesting that proportions of people who are gender non-conforming do not differ per country.

Effect of Gender Identification on Social Well-Being

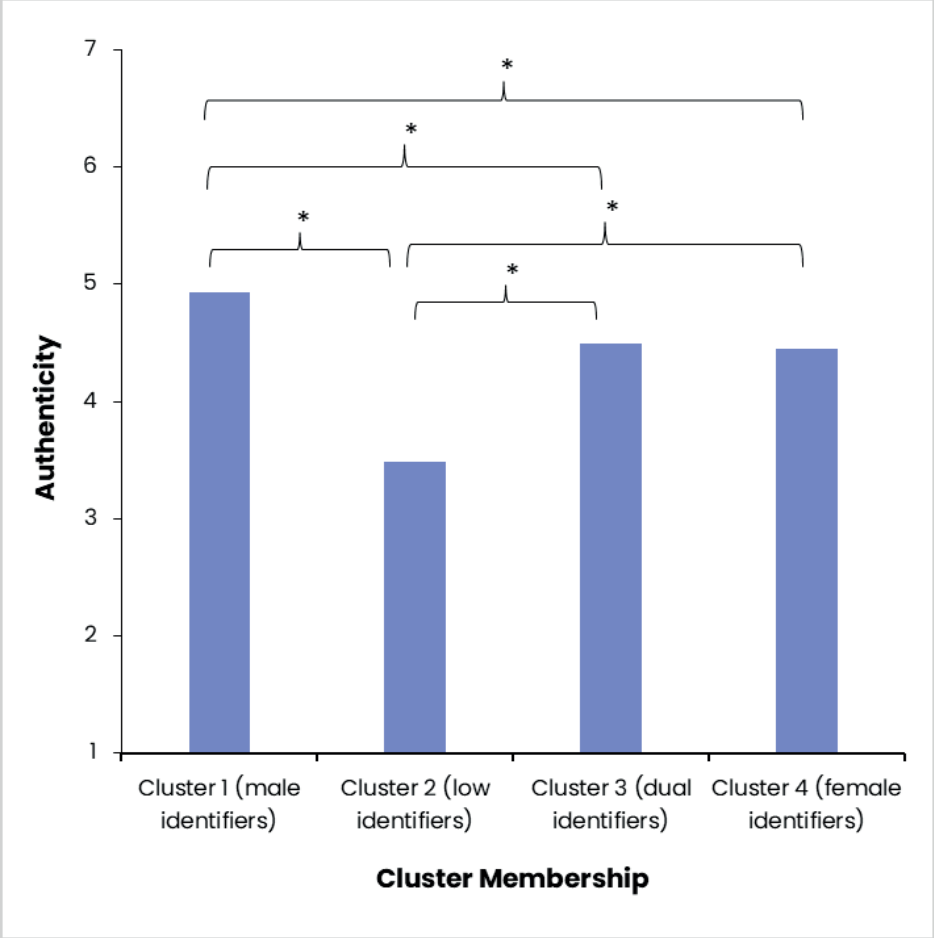
We wanted to test whether GI cluster membership is related to a person's social well-being, and whether this is moderated by self-assigned gender category and/or ASAB. We therefore ran a MANOVA with GI cluster membership (4 levels) as IV, self-assigned gender category and ASAB as moderators, and the following DVs: mean feelings of authenticity and belongingness in society, self-esteem, life satisfaction, and positive and negative affect towards answering previous questions.

As expected, we found a main effect of GI cluster membership on all measures of social well-being, except negative affect, which did not fit our prediction; authenticity, $F(3, 478) = 20.96$, $p < .01$, $\eta_p^2 = 0.12$; belongingness, $F(3, 478) = 13.21$, $p < .01$, $\eta_p^2 = 0.08$; self-esteem, $F(3, 478) = 6.18$, $p < .01$, $\eta_p^2 = 0.04$; life satisfaction, $F(3, 478) = 6.05$, $p < .01$, $\eta_p^2 = 0.04$; positive affect, $F(3, 478) = 13.23$, $p < .01$, $\eta_p^2 = 0.08$; negative affect, $F(3, 478) = 2.56$, $p = .06$). However, we did not find a main, or interaction effects, for self-assigned gender category, nor for ASAB (all *p*s > .05), contrary to expectation. To investigate which clusters differed from each other in terms of the different measures of social well-being, we ran Tukey HSD tests, which report separately for each measure, below.

Authenticity. The Tukey HSD test indicated that people in cluster 1 (male identifiers; $M = 4.93$, $SD = 1.29$) felt significantly more authentic (all *ps* ≤ .026) than people in all other clusters (2, 3 and 4). People in cluster 2 (low identifiers; $M = 3.49$; $SD = 1.52$) felt significantly less authentic than people in all other clusters (1, 3 and 4). Clusters 3 (dual identifiers; $M = 4.49$, $SD = 1.24$) and 4 (female identifiers; $M = 4.45$, $SD = 1.47$) did not differ in authenticity (see Figure 3).

Figure 3

Mean Feelings of Authenticity in Society per Gender Identification Cluster in Study 2

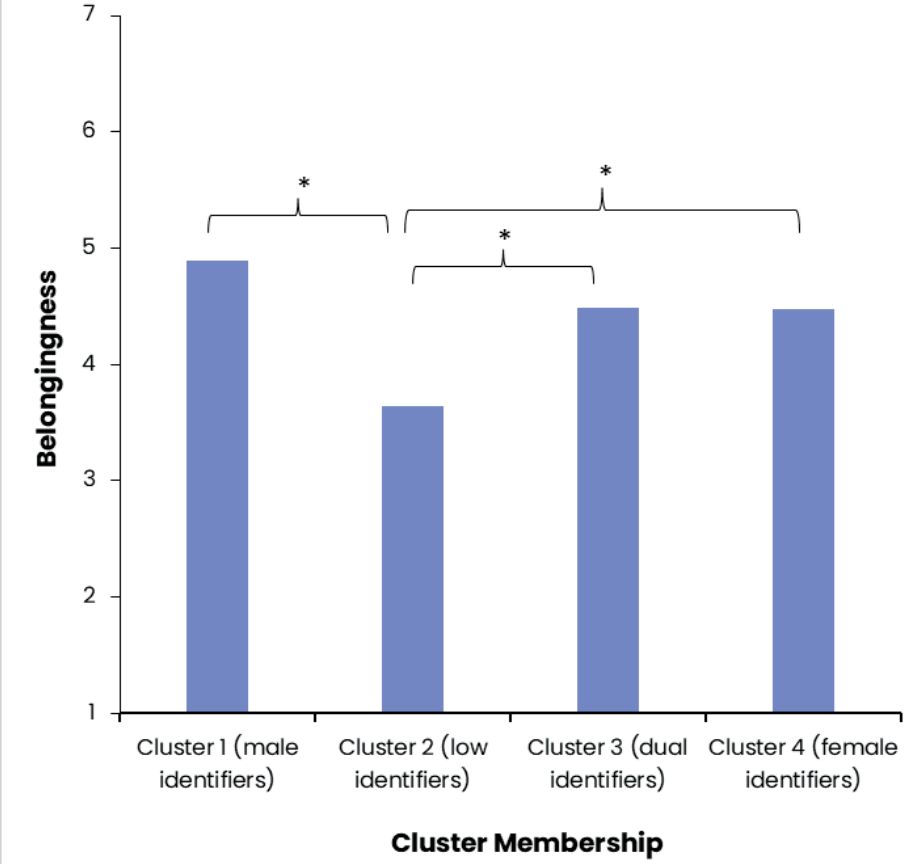


Note. Stars (*) denote significant differences between clusters ($p < .05$).

Belongingness. The Tukey HSD test indicated that people in cluster 2 (low identifiers; $M = 3.64$, $SD = 1.62$) felt significantly less societal belongingness (all $ps < .001$) than people in all other clusters (1, 3 and 4). Clusters 1 (male identifiers; $M = 4.89$, $SD = 1.36$), 3 (dual identifiers; $M = 4.49$, $SD = 1.35$) and 4 (female identifiers; $M = 4.48$, $SD = 1.61$) did not differ from each other in societal belongingness (see Figure 4).

Figure 4

Mean Feelings of Belongingness in Society per Gender Identification Cluster in Study 2

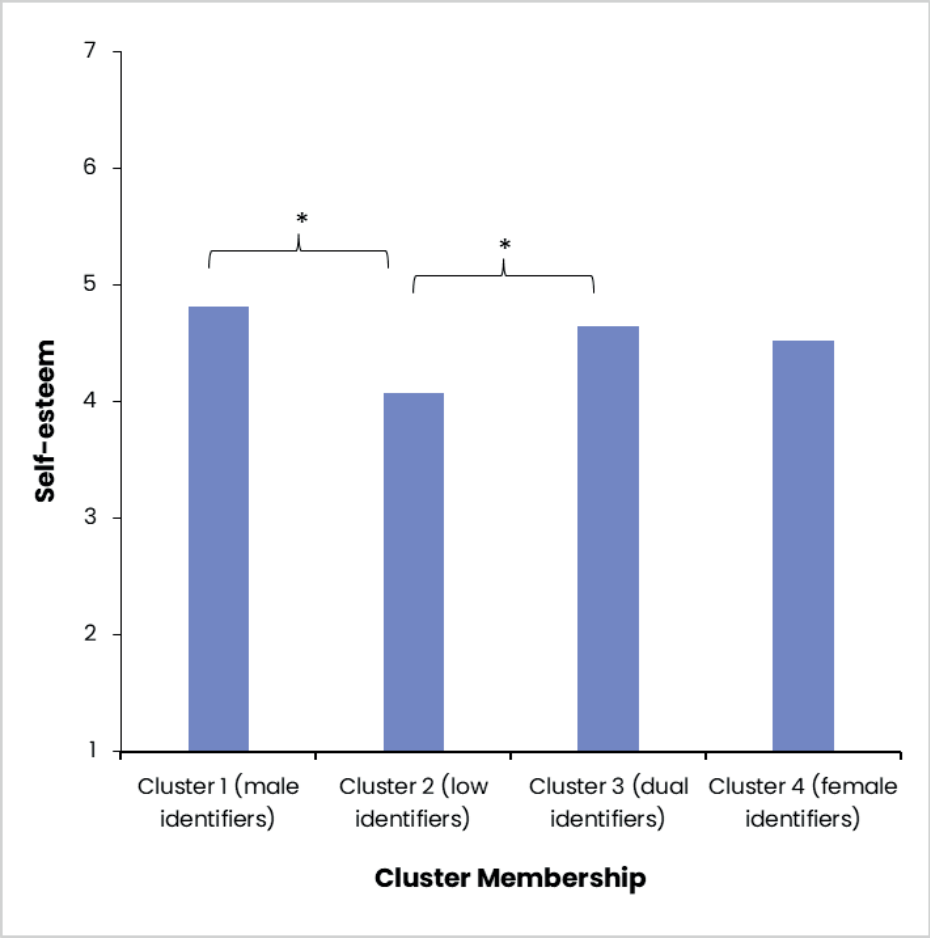


Note. Stars (*) denote significant differences between clusters ($p < .05$).

Self-Esteem. The Tukey HSD test indicated that people in cluster 2 (low identifiers; $M = 4.07$, $SD = 1.43$) felt significantly lower self-esteem (all $ps \leq .005$) than people in clusters 1 (male identifiers; $M = 4.82$, $SD = 1.39$) and 3 (dual identifiers; $M = 4.65$, $SD = 1.23$), but not 4 (female identifiers; $M = 4.52$, $SD = 1.27$). Clusters 1 (male identifiers) and 3 (dual identifiers) did not differ from each other in self-esteem (see Figure 5).

Figure 5

Mean Feelings of Self-esteem per Gender Identification Cluster in Study 2

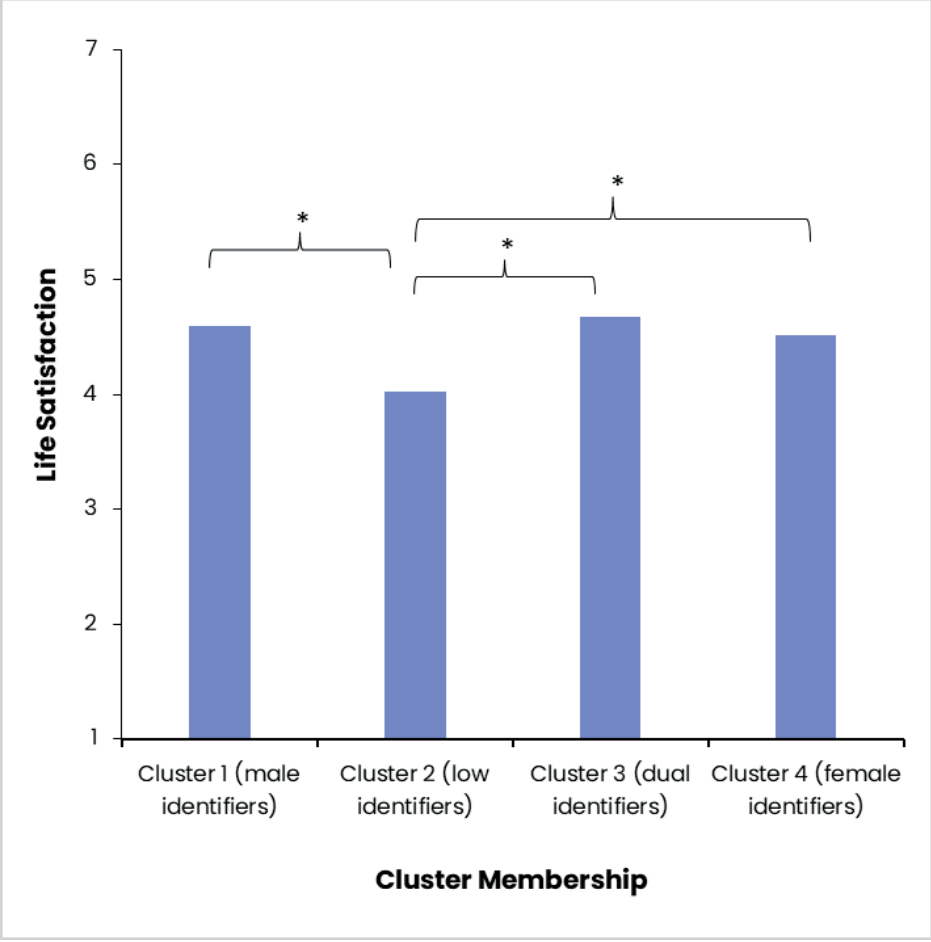


Note. Stars (*) denote significant differences between clusters ($p < .05$).

Life Satisfaction. The Tukey HSD test indicated that people in cluster 2 (low identifiers; $M = 4.03$, $SD = 1.31$) felt significantly less satisfied with their lives (all $ps \leq .020$) than people in all other clusters (1, 3 and 4). Clusters 1 (male identifiers; $M = 4.59$, $SD = 1.26$), 3 (dual identifiers; $M = 4.67$, $SD = 1.06$) and 4 (female identifiers; $M = 4.51$, $SD = 1.16$) did not differ from each other in life satisfaction (see Figure 6).

Figure 6

Mean Feelings of Life Satisfaction per Gender Identification Cluster in Study 2

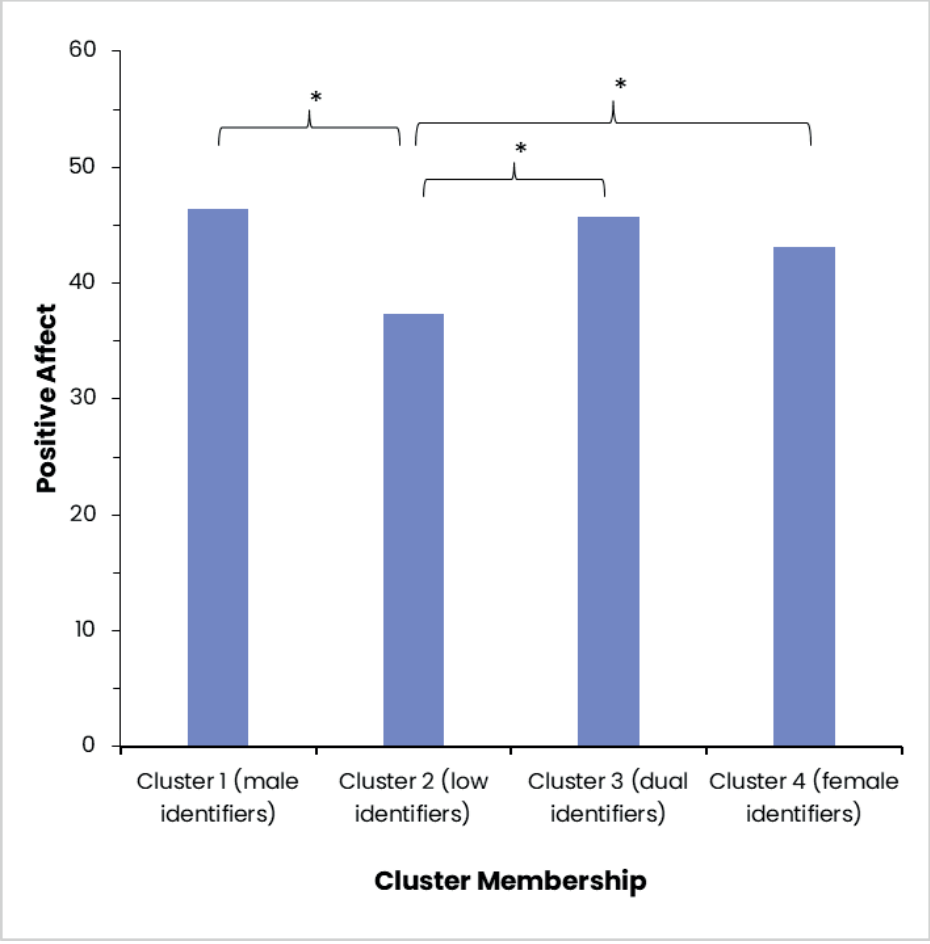


Note. Stars (*) denote significant differences between clusters ($p < .05$).

Positive Affect. The Tukey HSD test indicated that people in cluster 2 (low identifiers; $M = 37.37$, $SD = 13.14$) felt significantly lower positive affect (all $ps \leq .002$) than people in all other clusters (1, 3 and 4). Clusters 1 (male identifiers; $M = 46.36$, $SD = 11.19$), 3 (dual identifiers; $M = 45.77$, $SD = 11.12$) and 4 (female identifiers; $M = 47.07$, $SD = 11.44$) did not differ from each other in positive affect (see Figure 7).

Figure 7

Mean Feelings of Positive Affect per Gender Identification Cluster in Study 2



Note. Stars (*) denote significant differences between clusters ($p < .05$).

To see whether our measures of social well-being were related to one another, we ran a Pearson's correlation and found that all above measures were significantly correlated with one another (see Table 3).

Table 3

Correlations *between Social Well-being Measures, which We Found to Differ between Clusters*

Measure	Authenticity	Belonging-ness	Self-Esteem	Life Satisfaction
Belongingness	$r(480) = .89$ $p < .001$			
Self-Esteem	$r(480) = .59$ $p < .001$	$r(480) = .50$ $p < .001$		
Life Satisfaction	$r(480) = .59$ $p < .001$	$r(480) = .52$ $p < .001$	$r(480) = .71$ $p < .001$	
Positive Affect	$r(480) = .48$ $p < .001$	$r(480) = .42$ $p < .001$	$r(480) = .52$ $p < .001$	$r(480) = .50$ $p < .001$

Differences in Gender Identification Depending on Assigned Sex at Birth

To explore the relationship between ASAB and GI cluster membership, we ran a Chi-square test with ASAB as IV and cluster membership as DV. It revealed a significant difference in cluster memberships ($X^2 = 194.23$, $p < .01$) across AMABs and AFABs. A post-hoc z-score test revealed that the differences driving the significant effect were those in clusters 1, 3 and 4: AMABs more often fell into cluster 1 (male identifiers) than AFABs, AFABs more often fell into cluster 3 (dual identifiers) than AMABs, and AFABs more often fell into cluster 4 (female identifiers) than AMABs (see Table 4). Thus, while ASAB did not have a significant main effect, nor an interaction effect, on social well-

being, ASAB is related to cluster membership, which is in turn related to social well-being.

Table 4

Percentages of People Assigned Male or Female at Birth in each Cluster

Gender assigned at birth	Cluster 1 (male identifiers)	Cluster 2 (low identifiers)	Cluster 3 (dual identifiers)	Cluster 4 (female identifiers)
Male	48.35%	22.73%	25.21%	3.72%
Female	2.08%	17.08%	35.00%	45.83%

Discussion

We succeeded in our main goals for Study 2: a) to replicate the finding that a binary model is not the best fit for GI and report percentages of participants in each cluster and b) to test whether gender non-conforming participants report lower feelings of social well-being. While in Study 1 we had an overrepresentation of female participants, in Study 2 we recruited equal numbers of assigned male at birth (AMAB) and assigned female at birth (AFAB) participants. We therefore reported the cluster centres and percentages of participants in each of our four clusters. As suspected, we found a larger percentage of gender non-conforming individuals (approx. 50%) than previous work (3-5%; Kuyper & Wijsen, 2014; Van Caenegem et al., 2015), suggesting that gender non-conformity is more common than previously thought. Whereas in Study 1 we found that a three-cluster model was the best fit, in Study 2 we found that a four-cluster model was the best fit. A four-cluster model is in line with previous research done in children (Martin et al., 2017) and indeed in Study 2 we find the same four clusters as previous research. It seems that in Study 1, perhaps due to our sample

not having an equal sex distribution, we did not find enough participants who are low identifiers of both binary genders (cluster 2 in Study 2) as that was the missing cluster.

In Study 2, we also found that participants' GI was related to their feelings of social well-being, namely authenticity, belongingness, self-esteem, life satisfaction and positive affect, but not directly to their self-assigned gender category or assigned sex at birth (ASAB). Specifically, participants in cluster 2 (low identifiers), who are gender non-conforming, consistently reported the lowest social well-being on all measures, and also lower well-being than the other gender non-conforming cluster (3, dual identifiers). This is in line with previous findings of Martin et al. (2017) and Pauletti et al. (2017), who found that low identifiers experienced negative consequences to social adjustment. There is also extensive literature suggesting that lack of group identifications can have negative implications on health and well-being (Jetten et al., 2012), which may provide an explanation for this finding. Additionally, participants in cluster 3 (dual identifiers) reported being less able to be their authentic selves, an important part of social inclusion (Jansen et al., 2014), than participants in cluster 1 (male identifiers). However, there were no significant differences between participants in cluster 3 (dual identifiers) and participants in cluster 4 (female identifiers) on any of the well-being measures. This suggests that being a dual identifier is not related to additional negative consequences to social well-being, compared with being a female identifier. This may be because dual identifiers are able to socially adjust quite well, in a similar way as psychologically androgynous people (Bem & Lewis, 1975) have been found to be well-adjusted because of their sex role adaptability. Moreover, our results suggest that being a male identifier is related to the highest social well-being, as compared to all other gender identifications.

Interestingly, while there were no significant main or interaction effects of ASAB and self-assigned gender category on social well-being, we did find that ASAB was related to cluster membership (see below), suggesting an indirect effect on social well-being.

Lastly, we found that the distribution of people with different ASABs was different for each GI cluster membership. More AMABs were male identifiers than AFABs and more AFABs were female identifiers than AMABs. This seems to show that being brought up as a girl or boy increases identification with women and men respectively. Furthermore, we found that more AFABs than AMABs were gender non-conforming (specifically, dual identifiers) suggesting that high identification with both binary genders is more common among people who are considered women. This could be because women are the lower status group, which may lead them to want to identify more with men. In comparison, men may be motivated not to identify with women because they are the lower status group (e.g., Blau & Kahn, 2007), but we can only speculate about this.

A limitation of Studies 1 and 2 was that, while being less restrictive in that people could indicate their GI with more than one binary gender, we had only included items about the two binary genders: "man" and "woman." As such, a main aim of Study 3 was to test how participants would respond to a multiple-identification questionnaire that included a third gender, which we investigated in Study 3. Furthermore, in Studies 1 and 2, we had used the same questionnaire for measuring cognitive GI, which we had not yet compared to any other measure of self-categorisation. Our scalar measure relies heavily on participants' understanding of the English language, and there may be a concern that variance in our data is related to participants' different understandings of the items in the scale. As such, in Study 3, another main aim was to additionally include

a second (pictorial) measure of self-categorisation (as has also been done by Martin et al., 2017) to cross-validate the results of the scalar and pictorial measures, and to act as a conceptual replication of our results.

Study 3

Method

Participants

Based on a power analysis (effect size f = 0.20, smallest effect of GI on well-being from Study 2) we recruited 280 participants (140 AFAB, 140 AMAB; 140 identified as a man, 136 as a woman, 5 as a person with a third gender). Participants' age ranged from 18 to 63 (M = 36.97, SD = 8.58). Participants lived in 25 different countries, with the majority living in the UK (20.28%) or Portugal (18.51%). For information about participant country of residence, education, employment, and special needs, see Supporting Information.

Materials

All items were measured on a 7-point Likert scale unless stated otherwise.

Gender Identification. We used the same GI items as in Study 2. However, in addition to administering items about men and items about women, we also administered items about a third gender (e.g. "I see myself as someone belonging to a third gender"). We chose the term *third gender*, rather than a term like *non-binary*, because it encompasses all gender identities other than "man" and "woman," while not being too negatively loaded by sensationalist media (items about men α = .94, items about women α = .95, items about third gender α = .92; see Appendix A).

Overlap of Self and Group. As a pictorial measure of self-categorisation, we administered three items (one about men, one about women, one about a third gender) asking participants to indicate which

image, out of seven, best represents who they are in relation to each group. The images were Venn diagrams with one circle representing "you" and the other representing a gender group, which went from being very far apart to overlapping entirely (Schubert & Otten, 2002).

Social Well-Being. We administered the same Inclusion Scale (authenticity α = .96, belongingness α = .93), Self-esteem Scale (α = .89) and Life Satisfaction Scale (α = .86), as in Study 2[9].

Procedure

Participants were recruited through Prolific, using restrictions to recruit equal numbers of people in terms of ASAB. They were redirected from Prolific to the survey's Qualtrics link and asked to answer all questions. First, they were shown a text explaining that people can feel like a gender besides man or woman, that for the purposes of this experiment we will call any other gender besides man or woman a "third gender," and assuring them that any feelings they may have regarding their gender were perfectly normal. Next they filled in the scales in the following order: GI scale, Overlap of Self and Group items, Inclusion Scale, Self-Esteem Scale and Life Satisfaction Scale, Precarious Manhood and Womanhood (the latter is beyond the scope of this paper, but can be found in the Supporting Information). Items within scales were administered in a random order. Lastly, they answered demographic questions, read a debriefing, and had a chance to retract their data from further analysis. Participants received payment for their participation.

Results

Gender Identification

As in Studies 2 and 3, we calculated GIW and GIM scores based on the means of the GI items about women and items about men, respectively. Additionally, we calculated GIT (gender identification with a

third gender) scores based on the means of the cognitive GI items about a third gender group. For means, standard deviations and correlations of GI, see Table 5. The correlation between GIM and GIW was much higher in this sample than in Studies 1 and 2. This suggests that adding a third category to the scale alters people's responses to the items about binary categories. Furthermore, while our sample identified about equally with men and women (at around scale midpoint), $t(280) = 1.13$, $p = .26$, the overall identification with a third gender was significantly lower than both GIM, $t(280) = 16.40$, $p < .001$, and GIW, $t(280) = 17.67$, $p < .001$.

Table 5

Descriptive Statistics and Correlations of GIW, GIM and GIT

	GIW	GIM	GIT
Descriptive statistics	$M = 4.33$	$M = 4.07$	$M = 1.90$
	$SD = 2.01$	$SD = 1.95$	$SD = 1.13$
Correlation with GIW		$r(278) = -.90$*	$r(278) = .01$
		$p < .001$	$p = .86$
Correlation with GIM			$r(278) = .05$
			$p = .45$

Note. Absolute range of GIM, GIW, and GIT: 1-7.

Several two-step cluster analyses with three inputs (GIW, GIM and GIT) revealed that a 4-cluster model (as compared to 2, 3, 5 and 6-cluster models) was the best fit in terms of percentage of variance, Bayesian Information Criterion (*BIC* = 298.48) and silhouette of cohesion (silhouette ≥ .5) and separation combined. The cluster membership of each participant was saved to perform further analyses. The clusters can be interpreted as: 1) male identifiers, 2) multiple identifiers with male tendencies, 3) multiple identifiers with female tendencies and 4) female identifiers (see Table 6 for cluster centers and percentages of the sample in each cluster). This model differs from that of Study 2, in that there is no cluster that identifies lowly with both men and women.

Again, it seems that adding a third gender category altered the way participants thought about the items about binary categories. For the purposes of replication, we also ran a cluster analysis with two inputs (GIM and GIW), as well as a cluster analysis with the OSG measure, and found that a three-cluster model was the best fit for our data. This shows that a binary model was not the best fit, even when only including GIM and GIW as inputs. The results of this are beyond the scope of this paper but can be found in the Supporting Information.

Table 6

Cluster Centres for GIW, GIM, and GIT, and Percentage of Participants Who Fell into each Cluster

	Cluster 1 (male identifiers)	Cluster 2 (multiple identifiers with male tendencies)	Cluster 3 (multiple identifiers with female tendencies)	Cluster 4 (female identifiers)
Cluster centre for GIW	2.27	3.41	5.82	6.35
Cluster centre for GIM	5.97	5.00	3.14	1.97
Cluster centre for GIT	1.33	3.64	2.92	1.24
Percentage of overall sample	35.59%	17.08%	13.17%	34.16%

Note. Absolute range of GIM, GIW, and GIT: 1-7.

Once again, we investigated whether gender cluster membership differed depending on country of residence, as one might expect gender and culture to interact. As in Study 2, we found a non-significant result in a chi square testing this relationship ($X^2 = 82.56$, $p = 0.19$).

Overlap of Self and Group

To investigate whether the GI scale used in Studies 1 and 2 measured self-categorisation in a similar way that other measures do, we

compared our GI means with the Overlap of Self and Group measure. For means, standard deviations and correlations of the Overlap measure, see Table 7. We performed a Pearson's correlation analysis to compare GIM, GIW and GIT with each corresponding Overlap question, respectively. We found a significant and medium-large correlation between GIW and the Overlap of Self with Women measure ($r(278) = .77$, $p < .001$), between GIM and the Overlap of Self with Men measure ($r(278) = .81$, $p < .001$), and GIT and the Overlap of Self with Third gender measure ($r(278) = .71$, $p < .001$). All correlations were positive, medium-large ($r < .7$) and significant. We also looked at the mean Overlap measure per cluster membership and found that the patterns corresponded to those seen in the cluster centres for the GI measure (see Table 8). This suggests that while not measuring the exact same construct, the two measures overlap greatly.

Table 7

Descriptive Statistics and Correlations of the Overlap of Self with Women, the Overlap of Self with Men, and the Overlap of Self with Third Gender

	Overlap of Self with Women	Overlap of Self with Men	Overlap of Self with Third gender
Descriptive statistics	$M = 4.38$ $SD = 2.09$	$M = 4.13$ $SD = 2.14$	$M = 1.83$ $SD = 1.17$
Correlation with Overlap of Self with Women		$r(278) = -.64$* $p < .001$	$r(278) = .08$ $p = .19$
Correlation with Overlap of Self with Men			$r(278) = .22$ $p = .71$

Note. Absolute range of each Overlap measure: 1-7.

Table 8

Descriptive Statistics for Overlap of Self and Group per Cluster, M(SD)

Measure	Cluster 1 (male identifiers)	Cluster 2 (multiple identifiers with male tendencies)	Cluster 3 (multiple identifiers with female tendencies)	Cluster 4 (female identifiers)
Overlap of Self with Women	2.68 (1.38)	3.72 (1.70)	5.19 (1.43)	6.17 (1.35)
Overlap of Self with Men	5.96 (1.37)	4.81 (1.91)	3.38 (1.34)	2.17 (1.09)
Overlap of Self with Third gender	1.36 (0.76)	2.83 (1.58)	2.84 (1.01)	1.43 (0.72)

Note. Absolute range of each Overlap measure: 1-7.

Effect of Gender Identification on Social Well-Being

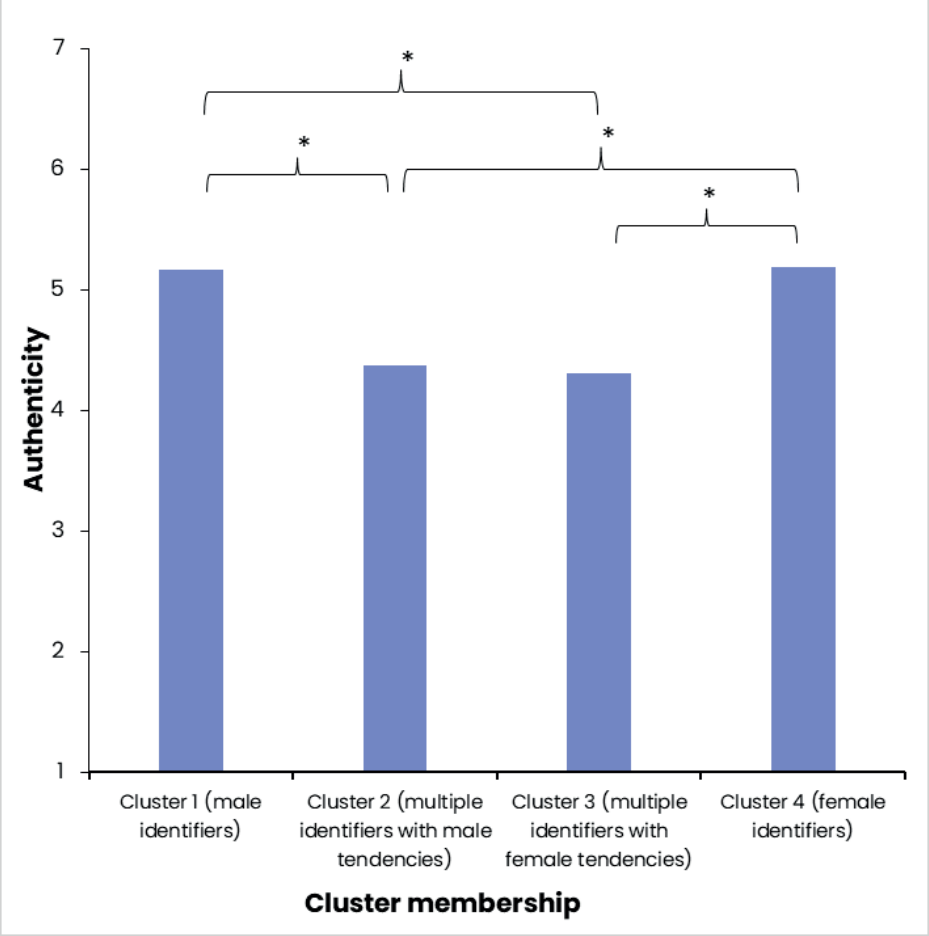
To investigate whether we can replicate the findings of Studies 1 and 2, that GI cluster membership is related to a person's social well-being, we ran a MANOVA with GI cluster membership (4 levels) as IV and the following DVs: mean feelings of authenticity and belongingness in society, self-esteem and life satisfaction. We found a main effect of GI cluster membership on all measures of social well-being; belongingness, $F(3, 277) = 6.92$, $p < .01$, $\eta_p^2 = 0.07$; authenticity, $F(3, 277) = 7.72$, $p < .01$, $\eta_p^2 = 0.08$; self-esteem, $F(3, 277) = 4.39$, $p < .01$, $\eta_p^2 = 0.05$; except for life satisfaction, $p = .056$. We once again explored the effects of ASAB and self-assigned gender category on social well-being (using a MANOVA), and whether these variables are moderators for the relationship between GI and social well-being. Unlike in Study 2, we found a significant main effect of ASAB on life satisfaction, $F(1, 267) = 7.37$, $p = .04$, $\eta_p^2 = 0.01$, such

that AMABs reported higher life satisfaction than AFABs, M = 4.30, SD = 1.35, M = 4.21, SD = 1.27, respectively. However, replicating Study 2, we found no other main or interaction effects. To investigate which clusters differed from each other in terms of the different measures of social well-being, which we had found a significant main effect of cluster membership for, we ran Tukey HSD tests, which report separately for each measure, below.

Authenticity. The Tukey HSD revealed that people in clusters 2 (M = 4.38, SD = 1.49) and 3 (M = 4.31, SD = 1.38; gender non-conforming clusters) reported significantly lower feelings of authenticity ($ps \leq .005$) than people in clusters 1 (M = 5.17, SD = 1.29) and 4 (M = 5.19, SD = 1.28; binary clusters) but did not significantly differ from each other. Clusters 1 and 4 also did not significantly differ from each other in terms of authenticity (see Figure 8).

Figure 8

Mean Feelings of Authenticity in Society per Gender Identification Cluster in Study 3



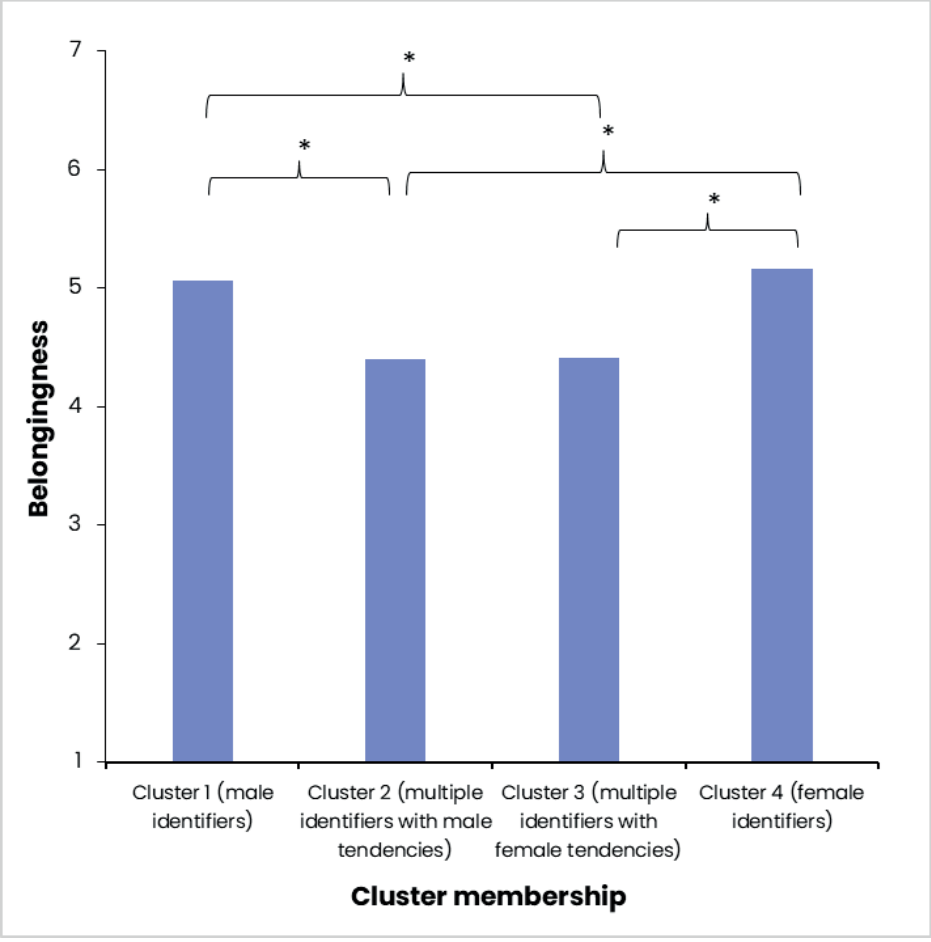
Note. Stars (*) denote significant differences between clusters ($p < .05$).

Belongingness. The Tukey HSD revealed that people in clusters 2 ($M = 4.40$, $SD = 1.26$) and 3 ($M = 4.41$, $SD = 1.29$; gender non-conforming clusters) reported significantly lower feelings of belonging ($ps \leq .026$) than people in clusters 1 ($M = 5.06$, $SD = 1.15$) and 4 ($M = 5.16$, $SD = 1.19$; binary clusters) but did not significantly differ from each

other. Clusters 1 and 4 also did not significantly differ from each other in terms of belonging (see Figure 9).

Figure 9

Mean Feelings of Belongingness in Society per Gender Identification Cluster in Study 3



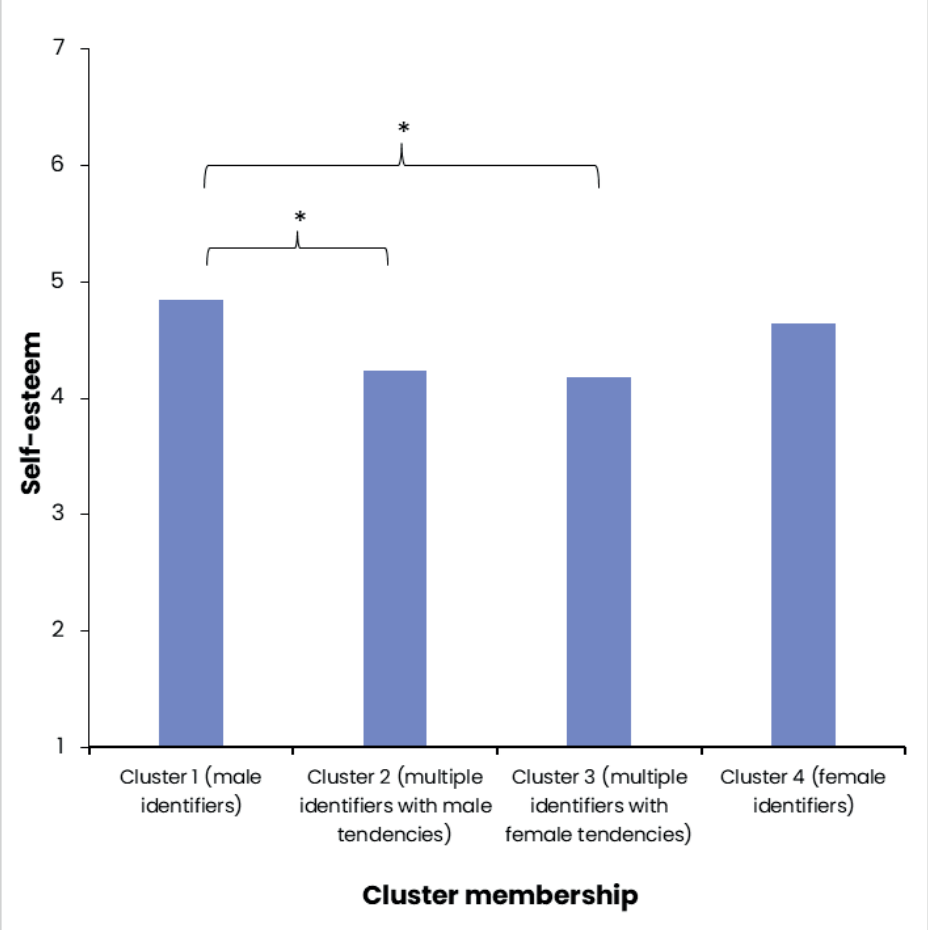
Note. Stars (*) denote significant differences between clusters ($p < .05$).

Self-Esteem. The Tukey HSD post-hoc test revealed that people in cluster 1 (male identifiers; $M = 4.84$, $SD = 1.08$) reported significantly higher self-esteem ($ps \leq .022$) than people in clusters 2 ($M = 4.24$, $SD = 1.26$) and 3 ($M = 4.18$, $SD = 1.28$; gender non-conforming clusters), but not people in

cluster 4 (M = 4.64, SD = 1.24). No other differences were significant, but the (non-significant) pattern of results follows that of the other variables of social well-being (see Figure 10).

Figure 10

Mean Feelings of Self-esteem per Gender Identification Cluster in Study 3



Note. Stars (*) denote significant differences between clusters (p < .05).

To see whether our measures of social well-being were related to one another, we ran a Pearson's correlation and found that all above measures, as well as life satisfaction, were significantly correlated with one another (see Table 9).

Table 9

Correlations between Social Well-being Measures

Measure	Authenticity	Belongingness	Self-Esteem
Belongingness	$r(278) = .83$		
	$p < .001$		
Self-Esteem	$r(278) = .41$	$r(278) = .47$	
	$p < .001$	$p < .001$	
Life Satisfaction	$r(278) = .46$	$r(278) = .51$	$r(278) = .68$
	$p < .001$	$p < .001$	$p < .001$

Differences in Gender Identification Depending on Assigned Sex at Birth

To explore the relationship between ASAB and GI cluster membership, we ran a Chi-square test with ASAB as IV and cluster membership as DV. It revealed a significant difference in cluster memberships ($X^2 = 218.60$, $p < .001$) across AMABs and AFABs (see Table 10). Post-hoc z-scores revealed that the proportions of AFAB and AMAB participants in all clusters differed significantly ($p \leq .05$), with more AMABs falling into clusters 1 (male identifiers) and 2 (multiple identifiers with male tendencies), and more AFABs falling into clusters 3 (multiple identifiers with female tendencies) and 4 (female identifiers).

Table 10

Percentages of People Assigned Male or Female at Birth in each Cluster

Assigned sex at birth	Cluster 1 (male identifiers)	Cluster 2 (multiple identifiers, male tendencies)	Cluster 3 (multiple identifiers, female tendencies)	Cluster 4 (female identifiers)
Male	70.50%	25.90%	2.16%	1.44%
Female	1.43%	7.86%	24.29%	66.43%

Discussion

We succeeded in our main goals for Study 3: a) to investigate the effect of adding a third gender to our multiple-identification model, and b) to cross-validate our GI scale with a pictorial measure of Overlap of Group and Self, both of which are thought to measure self-categorisation as a social group member (Turner & Reynolds, 2011).

Adding a third gender category to our multiple-identification model in Study 3 affected the results in that our correlation between GIM and GIW was noticeably higher, while still negative, than in Studies 1 and 2. Furthermore, we did not find any low identifiers of binary genders in our model in Study 3, as opposed to Study 2. It seems that the addition of a third gender category may deter participants from picking the lower scale points for binary genders. Self-categorisation theory (Turner & Reynolds, 2011) poses that we categorise ourselves as a member of an in-group, as compared to an out-group. Redefining the gender groups that participants were answering questions about (going from two to three), could therefore have caused them to rethink which groups are an in-group and which an out-group. In this case, the clearer out-group might have been people with a third gender, which would

increase self-reported identification with the in-group, namely binary genders. Importantly, we did find two clusters of people who reported medium identification with a third gender; the percentages of people falling into those clusters were sizeable (30.25%). This suggests that a sizeable amount of the sample understood a feeling of having a third gender outside of the binary, at least to some extent, and were not "put off" by answering questions about this. Considering our sample had a majority of cisgender, heterosexual participants, this is a notable and novel finding. Only five people in our sample chose "person with a third gender" as their explicit self-assigned gender category, while many more self-categorised to a certain extent as a third gender, showing the difference in results when measuring gender on a scale rather than as a categorical variable.

We found that our GI measure, which is comprised of a mean score of several items rated on a 7-point Likert scale, correlated highly and positively with the Overlap measure, which uses a single pictorial item to assess how much overlap a person perceives between their own identity and that of a certain group. Furthermore, participants' cluster centres for the scalar GI measure were comparable to the mean Overlap score for each cluster. All in all, this suggests that the two measures capture similar constructs, and we would come to similar conclusions using either measure.

Furthermore, replicating Study 2, we investigated how participants in different gender identification clusters differed in terms of social well-being. We found that participants in cluster 2 and 3 (both gender non-conforming clusters) reported lower feelings of authenticity and belongingness, than participants in cluster 1 and 4 (both binary clusters). Furthermore, we found that participants in the gender non-conforming clusters reported having lower self-esteem than male identifiers (cluster 1) but did not differ from female identifiers (cluster

4). With regards to authenticity, this finding is in line with findings from Study 2, where we found that participants in gender non-conforming clusters reported lower authenticity than participants in binary clusters. For belongingness, our finding in Study 3 differs somewhat from that in Study 2, where we only found that participants in one of the gender non-conforming clusters reported lower self-esteem than all other clusters. We think this is to do with the difference in content of each cluster in Study 3, as compared to Study 2. As discussed, in Study 3 we found that participants generally did not report identifying lowly with both binary genders, since providing a third gender option seemed to affect participants' reported binary gender identification levels. In Study 2, the group with the lowest reported social well-being (including belongingness) were participants who identified relatively low with both binary genders (but for whom we did not know how much they identified with a third gender). Lastly, our findings regarding self-esteem in Study 3 are somewhat in line with our findings in Study 2. Specifically, in Study 2 we had found that male identifiers reported having higher self-esteem than female identifiers and participants in one of the gender non-conforming clusters (low identifiers). In Study 3 we thus replicate that both gender non-conforming and female identifiers have lower self-esteem than male identifiers.

Finally, replicating Study 2, we explored the distributions of AFAB and AMAB participants in each gender identification cluster. Our findings were in line with what one might expect, namely that more AFAB participants were in clusters that identified relatively highly with women, while more ASAB participants were in clusters that identified relatively highly with men.

General Discussion

Across three empirical studies, we shed new light on gender non-conformity and its relationship with social well-being. The main

aim across all studies was to investigate the prevalence of adult gender non-conformity in a novel way: using social identification questionnaires and a cluster analysis inspired by Martin et al. (2017). We found the same result across all studies: that a binary model of gender is not as good a fit for our data as a multiple-identification model of gender. Using our multiple-identification models, we found a larger percentage of gender non-conformity in the general population than some of the previous literature (Kuyper & Wijsen, 2014; Van Caenegem et al., 2015). This suggests that gender non-conformity is a more common occurrence than previously thought. We argue that the questionnaires we provided participants gave them more space to express how they feel about their gender than in previous studies (e.g., Joel et al., 2014). This was due to the number of items used being higher and more diverse in terms of facets of gender that were addressed.

Gender Non-Conformity in Each of Our Studies

A discrepancy between results in the studies are the *number* of clusters found in the multiple-identification model: three in Study 1, four in Studies 2 and 3. We argue that because the sample in Studies 2 and 3 were more representative of the general World population in terms of ASAB (approx. 50/50), results from Studies 2 and 3 are likely closer to the true amount of gender non-conformity in the general population.

We also found differences in the *content* of the clusters between studies. Notably, in Study 2 we found a cluster of people who identified lowly with both binary genders, while no such cluster was found in Study 3. This was likely because of a difference in design: in Study 3 we measured GI with three gender groups, rather than two, which we had not done in previous studies. This difference in results between Studies 2 and 3 can be explained using Self-categorisation theory (Turner & Reynolds, 2011). Self-categorisation theory states that intergroup identification (which we measured) is the amount to which

one categorises with one group as compared to another group. The comparison of groups therefore makes a difference in self-reported identification. Comparing men, women, and people with a third gender, may make the comparison between binary and non-binary gender groups more salient than just comparing men with women.

Moreover, while one might expect a difference in gender identification between people of different countries because gender is culturally constructed (Newman, 1995), our results do not suggest that this is the case. In Studies 2 and 3 we found non-significant chi square results for the relationship between gender cluster membership and country of residence. We refrained from performing this analysis in Study 1 because of our less diverse sample, with most of our participants being from majority White, Westernized countries. In our studies, we made a start in investigating the effect of culture on gender identification, and our sample was multi-national in each of the three studies. However, the relationship between gender identification and country of residence can be investigated further, by reducing the White, Westernized bias in sampling further, and performing a more direct comparison between countries with different gender ideologies.

How Gender Non-Conformity and Social Well-Being Are Related

Adding to Study 1, in Studies 2 and 3 we examined whether and how GI relates to social well-being. Importantly, across both studies we found that gender non-conforming people consistently reported feeling less able to be their authentic selves in society, as compared to binary identifiers, showing that our multiple-identification model has real consequences. Thus, even though across both studies we had different methods, as well as different findings regarding gender identification clusters, we found that authenticity seemed to play a big role for gender non-conforming participants in both studies. A meta-analysis by Sutton (2020) showed that authenticity is consistently positively related to

important well-being outcomes, thus further highlighting the need for people to be able to express their authentic selves. This is an important indicator that acceptance of gender non-conformity in society needs to be addressed, so that they feel more able to be authentic.

Furthermore, there were some social well-being findings that were specific to each study. In Study 2, we found that low identifiers of gender (cluster 2) reported significantly lower well-being (all measures) than participants in all other clusters, including dual identifiers (cluster 3, i.e., the other gender non-conforming cluster). Identification with any group tends to be related to increased psychological well-being (e.g., Brook et al., 2008), so reporting no or low identification with all groups in a survey would likely cause participants to report lower well-being than those who do identify with (a) group(s). On the other hand, dual identifiers (cluster 3) in Study 2 seemed to be relatively well-adjusted, sometimes even better adjusted than female identifiers (cluster 4), in terms of social well-being (except authenticity). This finding is in line with previous research in children (Martin et al., 2017) and adults (Andrews et al., 2019, Endendijk et al., 2019) who found negative effects of low identifications and positive effects of dual identifications on various measures of social well-being (e.g., self-esteem, efficacy). This finding is also in line with research from Bem and Lewis (1975), who found that psychologically androgynous people, who report similar (relatively high) amounts of masculine and feminine traits, tend to be more adaptable and therefore have higher feelings of well-being. However, it should be noted that gender identification is a different concept from gender roles (measured through personality traits), as also pointed out by Martin et al. (2017) and Pauletti et al. (2017). One can be psychologically androgynous but identify highly with one binary gender (for more information about how our measure of gender identification is related to scores on the BSRI, see Supporting Information). Being adaptable

in terms of personality traits may be beneficial but feeling excluded by a society that does not accept feelings about gender beyond the binary may take its toll on people's well-being. In line with this, in Study 3, we found that participants in both gender non-conforming clusters reported lower feelings of belongingness in society than binary identifiers, as well as lower self-esteem than binary male identifiers. Therefore, even though gender non-conforming participants in Study 3 were relatively high identifiers of several genders, they reported suffering on several measures of social well-being. This further highlights the potential differences between psychological androgyny and gender non-conformity in terms of identification.

In both Studies 2 and 3, we also explored how ASAB and self-assigned gender category were related to social well-being, and how they affected the relationship between GI and social well-being. We theorized that ASAB and self-assigned gender category are indicators of how one is perceived by others, and that this could influence social well-being. For instance, someone who is assigned male at birth is typically raised as a boy and is often perceived by others as a boy or man throughout his life. Men tend to be afforded more privileges than women (Eagly, 1983) and therefore may feel more included by society in general. In Study 2, we found no significant results of ASAB or self-assigned gender category (neither main nor interaction effects) on social well-being. In Study 3, we largely replicated these results, with one exception. In Study 3, we found a significant main effect of ASAB on life satisfaction, such that AMABs reported higher life satisfaction than AFABs. This is in line with the idea that people who are mostly perceived as men may have more privilege than people who are mostly perceived as women, and therefore may feel more satisfied in their lives. What remains unclear, however, is why we found this result in Study 3, but not Study 2. What was consistent across both studies though, was that the

effect of gender identification on social well-being is not moderated by ASAB or self-assigned gender category. We believe this underlines the theoretical implications of our multiple-identification model, and that our self-categorisation measure can be related to real psychological consequences, without it necessarily acting as a proxy for other ways of measuring gender and sex.

Limitations

As stated, there was a discrepancy in our findings regarding the different clusters of GI whereby we found a different number of clusters, and different meanings attached to each cluster, across our three studies. This is likely to do with a lack of direct methodological replication in our studies: the measures in our survey, and the demographics of our sample, differed from study to study. Furthermore, the correlation of GIW with GIM showed a vastly different coefficient across studies. While in Studies 1 and 2, the correlation coefficients were relatively small, suggesting discriminant validity between the measures of GIM and GIW (Campbell, 1960), in Study 3, we found a high correlation coefficient between these two measures. This seems to show that, when asked to consider to what extent they identify with a third gender, participant responses shift such that GIM and GIW become highly related to one another. It should be noted that, even in Study 3, where GIM and GIW were highly related to one another, the best cluster solution for our data, with two inputs (GIM and GIW), was not binary but, rather, contained three clusters. As such, while the high correlation between GIM and GIW suggested that the two measures were on the same continuum (lack of discriminant validity), measuring both allowed us to find more than two clusters of gender identification. It is important for future researchers to investigate this finding further, and to run direct methodological replications of their studies.

In order to measure GI, we used a Social Identity Theory (Tajfel, 1974) approach taken from Ellemers et al. (1999). However, there are alternative ways to measure gender identification. Our approach could be seen as measuring the self-investment component of gender identification (i.e., "solidarity" with a group, "satisfaction" with a group, and identity "centrality"), while there is also the self-definition component of gender identification, which measures self-stereotyping and group homogeneity (Leach et al., 2008). Measuring self-definition was not part of the scope of this paper. However, it will be important in future studies to investigate how estimates of gender non-conformity differ depending on the way we measure their identification.

Next to the different ways of measuring identification with men and women, we also did not measure how our results would change if we had used different gender terminology than "man" and "woman." Schudson et al. (2019) demonstrated that different gender words are associated with different definitions. While the words "male" and "female" tend to be associated with biological cues, such as body shape and genitals, the words "masculine" and "feminine" tend to be associated with social cues, such as clothing. Importantly, they found that the words "man" and "woman" are associated with a little bit of both types of gendered/sexed information: biological and social. Had we used words associated with just biology, for instance, we might have found that a binary model (two clusters) may have had the best statistical fit, as biological sex is seen as more fixed than gender (Skewes et al., 2018). On the other hand, had we used words associated with social information, for instance, we might have found that a three-cluster model would have been the best fit, similar to Bem (1974; "masculine," "feminine," and "androgynous" people).

Participants in our studies were recruited from the sample of the online platform Prolific. This sample is relatively representative of

the general population (Palan & Schitter, 2018). However, the majority of participants who are active on Prolific are White and British. In Study 1, we found a much higher percentage of White participants and participants living in the UK than in any other country. This is not ideal since the changes in the gender narrative are happening world-wide and not just in the UK. However, in Study 2 we had more participant requirements and therefore our sample was much more international and ethnically diverse, meaning the White British-centric bias of Study 1 was reduced. Since results of both studies mostly converge, we assume that the UK bias in Study 1 was not detrimental to our results. Furthermore, our measurements were all explicit and self-reported. In future research we aim to use more implicit measures of gender identification, to further grasp the prevalence and implications of gender non-conformity.

Finally, while the main scope of this paper was about strict gender categorisations that are in place in society, gender may not be the only social category that is falsely assumed to be binary. Social categories such as race (Black or White) or age (old or young) might behave in similar ways as the gender results described in this paper. It is worth noting that we made a start at comparing binary gender and binary ethnic categories and found that ethnic categories did not behave the same way as gender categories; the details of this can be found in the Supporting Information.

Practical Implications

We believe that the data reported in this paper speak to the current rise in gender-inclusive interventions (third gender markers in legal documents, e.g., Dunne & Mulder, 2018; gender-neutral toilets, e.g., Gershenson, 2010; unisex clothing, e.g., Park, 2014; gender-neutral language, e.g., Hord, 2016). These interventions are designed to include people who identify as "non-binary" (their self-assigned gender category is "non-binary"). However, we believe that these interventions could help

increase feelings of inclusion in anyone who self-categorises, in one way or another, in a gender non-conforming way. We have shown that the number of people who identify in a gender non-conforming way may be much larger than previously thought, and that gender non-conformity is related to social well-being, including how included or excluded people feel by society. We argue that an increase in gender-inclusive interventions may therefore be a positive change for many, including cisgender, heterosexual people who identify in a gender non-conforming way. However, more research is needed to draw a direct link between gender-inclusive interventions and people's higher social well-being, as well as how to design these interventions to maximize positive effects.

Conclusion

In this paper, we show that a binary model of gender ("man" vs. "woman") is limited. Across three studies, we found that a binary model does not fit the general population as well as a multiple-identification model of gender. This means that measuring gender in a binary way (women are only asked how much they identify with women, and vice versa) misses some of the complexities and nuances of people's gender identification. If we continue to do so, we are likely to draw a skewed picture in (social psychological) research. Furthermore, there are real consequences to identifying in gender non-conforming ways, as we found that individuals identifying in a gender non-conforming way often suffer from lower social well-being (particularly, feeling like they cannot be their authentic selves). Our research therefore reveals why maintaining a binary system in society may be harmful to a significant proportion of the population.

Appendix A: Three-way Gender Identification Scale

1. I identify with women
2. I identify with men
3. I identify with a third gender
4. I identify as a woman
5. I identify as a man
6. I identify as a person with a third gender
7. I am like women
8. I am like men
9. I am like people with a third gender
10. Women are an important reflection of who I am
11. Men are an important reflection of who I am
12. People with a third gender are an important reflection of who I am
13. I see myself as someone belonging to the group of women
14. I see myself as someone belonging to the group of men
15. I see myself as someone belonging to a third gender
16. I have a lot in common with women
17. I have a lot in common with men
18. I have a lot in common with people with a third gender

Footnotes

[1] For a more detailed explanation of our power analysis reasoning, please see Supporting Information.

[2] We checked for differences in cluster memberships depending on demographic profile by means of chi-square analyses and Pearson's correlations but found no significant results.

[3] We additionally asked participants about their emotional connection to each gender group, and their evaluation of each gender group, which are other subcomponents of gender identification, in line with Ellemers et al. (1999). There were several issues with these measures, such as cross-factor loadings, low reliability, and lack of replication between Study 1 and Study 2 (which is also why we excluded these measures from Study 3). Because of this, the results of these measures are beyond the scope of this paper but can be found in the Supporting Information.

[4] You can find the results of these questionnaires in the Supporting Information. The exceptions to this are the questions about feminine and masculine norms. We asked participants to fill out shortened versions of the Conformity to Feminine Norms Inventory (CFNI; Mahalik et al., 2005) and the Conformity to Masculine Norms Inventory (CMNI; Mahalik et al., 2003). The goal, initially, was to measure a general conformity to feminine or masculine norms. However, a recent study has shown that the CMNI does not reliably measure a general conformity to masculine norms and is only reliable for measuring conformity to specific norms (Hammer et al., 2018), which we were unable to do with our shortened versions of the questionnaire. We concluded that the CFNI is similarly problematic and therefore refrained from analysing data from either measure.

[5] We also ran separate analyses within male and female participants on the relationship between GIM and GIW. We found that the pattern

of results of male and female participants were similar (with one result being non-significant in male participants but significant in females, likely due to lower male sample size), meaning we come to the same conclusions regardless of the participant's gender. The details of these analyses can be found in the Supporting Information.

[6] For more information about the reasoning for our power analysis, please refer to Supporting Information.

[7] The word "ethnicity," rather than "race," was used in Prolific's pre-screening question at the time. Prolific have since changed their pre-screening question and use the word "race" instead, which more accurately reflects the answer options they provide (e.g., White) which are racial, rather than ethnic, categories. However, to stay consistent with the wording that was given to our participants, we have used the term "ethnicity" throughout this paper.

[8] We additionally asked participants about their emotional connection to each gender group, and their evaluation of each gender group, in line with Ellemers et al. (1999). Similarly as in Study 1, the results of this are beyond the scope of this paper but can be found in Supporting Information.

[9] We chose not to include the Positive And Negative Affect Schedule (PANAS), as in Study 2, because we were least interested in this measure. The other measures of social well-being were more easily generalizable to social well-being more broadly, which better fit the aim of Study 3.

Chapter 5: Exploring implicit associations of self- and other-concepts with gender

"Girls who are boys who like boys to be girls who do boys like they're girls, who do girls like they're boys"

Girls & Boys – Blur

Author contributions: Miriam I. Wickham (Conceptualisation; Data curation; Formal analysis; Investigation; Methodology; Validation; Writing – original draft; Writing – review & editing); Félice van Nunspeet (Conceptualisation; Methodology; Validation; Writing – review & editing); Naomi Ellemers (Conceptualisation; Funding acquisition; Methodology; Validation; Writing – review & editing)

Abstract

The gender binary is the pervasive view that people can only be men or women, and that these two gender groups are mutually exclusive. Conversely, previous research (Wickham et al., 2022) has shown that people do not explicitly see themselves as only male or female, but rather, as a mix of both. Across two empirical studies, we investigated how people view themselves in an implicit, rather than explicit, way. We additionally investigated people's implicit concepts of other people's genders and how those relate to their self-concepts. We conducted Study 1 (N =104) at a popular Dutch science festival, using a short version of a reaction time task. In Study 2 (N = 94), we replicated our previous findings among a larger sample, with a longer and more controlled experiment. We found that people's implicit self-concepts are also not strictly binary, replicating previous (explicit) results. For concepts of other people's genders, we found that people also see others as a mix of both male and female, and thus also not in a strictly binary way. Furthermore, we found a strong relationship between implicit gender self-concepts and implicit concepts of one's ingroup, but not one's outgroup, suggesting that self-categorisation was more linked to ingroup prototypicality than distinctiveness from the outgroup. Our findings add to existing literature about gender non-conformity, gender self-categorisation, and gender social cognition.

The World Health Organization recently updated its guidelines for health managers, to reflect that neither sex nor gender are binary (WHO, 2022). This is an example of a societal shift in how we view gender. Specifically, the societal view is shifting away from the gender binary, whereby everyone is seen as either male or female, towards a more inclusive view of gender as a spectrum. This may have consequences for both our self-concepts, that is, how we view our own genders, as well as our concepts of others and their genders. Previous research has shown that a large proportion of the population do not fit into a strictly binary model in terms of how they view their own gender (Wickham et al., 2022). This previous research was done using explicit measures of gender identification. In the present chapter, we sought to expand on previous research, by using an implicit measure of gender identification, which can provide insight into people's more unconscious and automatic gendered self-concepts. Furthermore, we sought to additionally investigate how people implicitly consider other people's genders and whether these are also not seen as strictly binary. This provided us with insight into implicit gender categorisations of others. Lastly, given that our self-concepts as members of social groups are tied to how we view those social groups, we explored how gendered self-concepts are related to concepts of others' genders.

Gender beyond the binary

In recent years, a societal debate surrounding gender has emerged, whereby people are increasingly questioning whether gender is binary (i.e., people fit into two gender categories: men and women) or should rather be considered a spectrum (Cooper, 2019). As a result of this, we have seen a rise in gender-inclusive initiatives that are aimed at improving the lives of those who do not fit into the gender binary (Hord, 2016; Pun et al., 2017; Reis et al., 2018). Similarly, some

social psychological researchers have argued that we must research gender from a non-binary perspective, so as to address this societal debate (Morgenroth & Ryan, 2018).

A special focus of previous research has been whether or not people's gender identifications, that is, their self-categorisations and feelings of belonging to a certain gender group, are binary (Abed et al., 2019; Andrews et al., 2019; Baiocco et al., 2022; Endendijk et al., 2019; Joel et al., 2014). A study by Wickham et al. (2022), for instance, found that, in a large international sample, participants' gender identifications consistently clustered into more than two categories, suggesting that a binary model of gender was not the best fit. Crucially, previous research has investigated gender identification using explicit measures (i.e., self-report) only. Implicit measures, on the other hand, may capture people's gender self-concepts in a different way and may thus give us further insight into whether people see themselves in a binary way or not. Specifically, implicit measures are argued to tap into more stable, unconscious and automatic processes (Gawronski & Payne, 2010; Payne et al., 2008), and are less prone to social desirability and ambiguity in the interpretation of language (Uhlmann et al., 2012). Furthermore, previous research has had a strong focus on gender beyond the binary in terms of how we categorise *ourselves*, while there has been less of a focus on how we categorise *others* (in a binary way or not). How we categorise others is a relevant topic to study, since categorising others into groups is at the core of stereotypes, prejudice and intergroup conflict (Krueger & DiDonato, 2008), and can thus give us insight into societal norms around gender. Furthermore, as we know from *self*-categorisation theory (Turner & Reynolds, 2011), our self-concepts are also tied to how we categorise *others*, highlighting the importance of the relationship between the two.

Implicit gender self-concepts

Researching gender self-concepts implicitly can give us further insight into whether people tend to view themselves in a binary way or not. Initially it was thought that implicit measures tap into more affective unconscious processes, and thus less cognitive or top-down, but this highly debated (Gawronski et al., 2017; Gawronski & Sritharan, 2010). Nonetheless, there are other reasons why including implicit measures is valuable, namely that they are less prone to social desirability and rely less on language comprehension. Specifically, participants are often less able to deliberately perform in a specific way during an implicit task, and researchers can thus tap into beliefs or processes that participants may not be aware of themselves. Furthermore, while explicit measures often rely on language comprehension (e.g., questionnaires with fixed answer scales), implicit measures often make use of less words and more pictures (Uhlmann et al., 2012). Thus, implicit measures can be applied across different groups (of different cultures or abilities) since they rely less on everyone interpreting and processing language in the same way.

Gender, in particular, is an excellent social category to study using an implicit measure, given that gendered language and symbols are so pervasive in society. Dutch (the language our studies were conducted in), just like most Indo-European languages (Corbett, 2013), is highly gendered in that we often cannot speak about a person without referring to their gender (for instance, by using pronouns). Similarly, in the Netherlands, as well as many other countries, we are faced with gendered icons on a daily basis (for instance, when deciding which bathroom to use; Enke, 2016). This makes (binary) gender a highly salient identity which is primed very regularly through gendered language and symbols which a vast majority of people are very familiar with. In our study, our aim was thus to use an implicit

measure with common words and symbols to tap into somewhat more unconscious ideas about how the self fits into these pervasive societal concepts. While this was exploratory, we expected that we might be able to replicate results from previous research using explicit measures (Wickham et al., 2022), whereby it was found that people's self-concepts did not fit a strictly binary model of gender. Replicating this finding in an implicit way would further show that the societal debate about the gender binary reflects that people do not see themselves in a binary way.

Gender self-concept and other-concept

In addition to people's implicit self-concepts, we were also interested in people's implicit concepts of other people. How we view ourselves is different from how we view others. Our self-concepts are more tied to our perceptions of our own sensations, emotions and cognitions, and thus a reality that we have more insight into than that of other people. Our concepts of others, on the other hand, are linked to our observations of others' behaviours and expressions, from which we assume to understand their realities (Pronin, 2008). To understand our own gender we tend to rely more on our own feelings and thoughts about gender, and process those internally, rather than externally (Levi, 2006). To understand someone else's gender, on the other hand, one would usually try to use perceptual information regarding their biological features (Mouchetant-Rostaing et al., 2000), how they are dressed (Leinbach & Fagot, 1993), or their feminine or masculine behaviour (Martin & Halverson, 1983).

However, our self-concepts and other-concepts are also linked, given that the way we perceive and form social groups is related to how we view ourselves as members of those social groups (Turner & Reynolds, 2011). For instance, level of group identification (how much one sees oneself as a member of a group) is related to how we deal

with threats to group distinctiveness (Spears et al., 1997). Specifically, when the distinction between groups (e.g., men and women) is at threat, people who identify highly with their group will find ways to deal with that. For example, they may self-stereotype, that is, perceive and describe themselves in line with stereotypes about their own group (such as women being nurturing), in order to increase distinction from the out-group. Thus, while our self-concepts are distinct from our concepts of others because of different underlying processes of each, they are also related to one another. As such, exploring both within the same study can help us understand how they are distinct, but related, in the specific case of gender beyond the binary. In our studies, we explored both implicit gender self-concepts and implicit concepts of others' (both in-group and out-group) genders.

The present studies

Across two studies, we measured participants' explicit gender identification, implicit gendered self-concepts and implicit concepts of others' genders. Specifically, we asked participants to complete a reaction time task designed to measure their implicit associations of themselves, women in general, and men in general, by asking them to respond to gendered words and icons. We compared their implicit self-concepts to an explicit measure of gender identification, to measure to what extent explicit and implicit gender self-concepts were related to one another. We also compared their implicit gender self-concepts to their implicit concepts of others' genders, to explore how these associations about the self were related to implicit associations about the ingroup or outgroup.

The two studies were very similar, with the second study being a conceptual replication and extension of the first. Study 1 was conducted at a popular Dutch science festival, which allowed us to swiftly and efficiently collect a lot of data while trying out a new

approach to measuring gender self-concepts. The purpose of Study 2 was to replicate our previous findings among a larger sample, while using a longer experiment which participants conducted in a more controlled environment.

Our experiment included gendered stimuli, for which we collected words and icons that are fundamentally gendered and which we come across in our daily lives, thus our participants would most likely be very familiar with them. While the words and icons we used in our reaction time task were binary (i.e., they were related to either men or women, and no other gender groups), we argue that our measure of implicit gender concepts is not binary. This is because we measured to what extent participants' self-concepts were male *as well as* female, regardless of their self-assigned gender identity. As such, we were able to measure to what extent people implicitly see themselves as a mix of both male and female. Conversely, a strictly binary measure would be measuring women's female self-concepts and men's male self-concepts *only*, as that would imply that women can only see themselves as female (and never male), and vice versa for men.

We had several predictions which we wanted to test. First, we expected to find a moderate relationship between participants' explicit gender identification and implicit gender self-concept, such that the two would converge but be somewhat distinct, in line with literature on the relationship between implicit and explicit processes more generally (Payne et al., 2008).

Second, we expected to find that participants' implicit self-concepts would not be binary, following previous findings using explicit measures. To establish this, we followed the analysis strategy of Wickham et al. (2022), whereby we tested to what extent participants may see themselves as *both* male and female, which would go against the binary idea that male and female are mutually exclusive.

Specifically, we tested whether implicit gender self-concepts (male and female) correlated relatively weakly, which would suggest that seeing oneself as male or as female are not polar opposites in people's implicit associations, following Wickham et al. (2022). Second, we tested the multiple identification model of gender (Wickham et al., 2022), such that we clustered participants according to their implicit male and female self-concepts, to see whether we would find more than two clusters. This would show that a binary model of implicit gender is not the best fit.

In a similar way, we explored whether participants' concepts of others' genders were binary but had no specific expectations about this. Specifically, we investigated whether implicit concepts of others' genders (male and female) would correlate weakly (for all predictions regarding correlation size, we followed guidelines by Cohen, 1992), suggesting that seeing others as male or as female were also not polar opposites in people's implicit associations. Lastly, we visually and statistically compared implicit gender self-concept and implicit concepts of others' genders and expected that we would find a relationship between them, but had no specific expectations regarding the nature of this relationship.

Study 1

Methods

Participants

Participants were visitors of the popular Dutch science festival "Betweter" in Utrecht, the Netherlands. We recruited 153 participants but had to exclude 49 participants due to missing data or disruptions during their participation. Our sample thus consisted of N = 104, of which 27 were self-identified men and 77 were self-identified women[1]. No participants had indicated taking any drugs, and most of our participants had indicated consuming no more than three

units of alcohol (90.38%), with a maximum of 6 units reported by one participant[2]. To minimise the duration of participation, we did not ask for any other demographics.

Materials

Implicit gender associations. To measure participants' implicit associations about women's and men's gender, as well as participants' own (binary) gender, we asked participants to do an Implicit Relational Assessment Procedure (IRAP) based on Cartwright et al. (2017). Participants were shown icons representing either women or men (e.g., the typical women's and men's bathroom signs), as well as words that are fundamentally gendered (e.g., the personal pronouns "she" and "he"[3]; for a full list of stimuli, please see Appendix A). Above each icon or word, they either saw "men =", "women =" or "I ="[4]. Participants were asked to respond to each trial with "true" or "false", by pressing the letters 'q' or 'p' on their keyboards (which of the letters represented "true" or "false" was randomized across participants, as such controlling for left-right effects), according to a rule (see Figure 1 for an example of a trial). In the gender-congruent block, the rule was that women were female (women = [typically female icon/word]) and men were male (men = [typically male icon/word]), and vice versa in gender-incongruent blocks. For female participants, the rule in the gender-congruent block was that participants themselves were female (I = [typically female icon/word]), while for male participants, the rule in the gender-congruent block was that participants themselves were male (I = [typically male icon/word]). If they responded incorrectly, they were shown a red 'X' for 500ms after the trial. The rule changed from block to block, such that there was one gender-congruent and one gender-incongruent block, as well as two practice blocks (one before each real block). The order of the blocks as well as the order of the icons and words were randomized, thus controlling for order effects. Before the first block, participants were shown all icons and images to

familiarize them to each pair of stimuli. Each block contained 24 trials. Participants were given a maximum of 2000ms to respond to each trial. After each trial, there was a break of 500ms, where participants viewed a fixation dot, before the next trial began. After each block, participants could rest for as long as they wanted before beginning the next block. By measuring the reaction times and accuracy of participants, we were able to calculate so-called *d* scores (Greenwald et al., 2003) to get an indication of how easy or difficult it is for participants to learn gender-congruent as compared to gender-incongruent rules. Specifically, *d* scores were calculated as the difference in reaction times between conditions, divided by a pooled *SD* of all correct trials. Error latencies were replaced with a replacement value ($M + 2\ SD^{correct}$) and replaced latencies exceeding the maximum response time with the maximum response time of 2000ms. With this IRAP, we extracted six d scores, for the implicit beliefs that 1) participants are female ("I am female"), 2) participants are male ("I am male"), 3) women are female, 4) men are male, 5) women are male, and 6) men are female.

Figure 1

Example of a Trial

Explicit gender identification. To get an indication of participants' level of gender identification, we asked to indicate how much they agreed, on a 7-point Likert scale (1 = "completely disagree", 7 = "completely agree"), with the following items: "I identify with men" and "I identify with women"[5] (Postmes et al., 2013).

Procedure

Participants were recruited among passers-by at "Betweter" festival and told that the experiment would take between three and five minutes. In order to minimise the effects of the noise and visual stimulation of many passers-by, participants completed the experiment behind a curtain, while wearing noise-cancelling headphones and

listening to white noise. They completed the experiment on Gorilla Experiment Builder (Anwyl-Irvine et al., 2020) and gave informed consent digitally. All participants completed the experiment in Dutch. They were asked to indicate how many units of alcohol they had consumed, and whether they had consumed any drugs. Next, they indicated their gender and gender identification. Next, they completed the implicit gender associations task. Lastly, they were debriefed and given the opportunity to ask any questions to the main experimenter.

Results and Discussion

Implicit associations of own gender

We extracted *d* scores, based on accuracy and reaction time, representing the implicit associations between the self ("I") and the two binary gender groups (male and female), such that we had scores representing two implicit beliefs: "I am male" and "I am female". We explored how these implicit measures compared to the explicit gender identification measure, by comparing the patterns of descriptive results and analysing the correlation between implicit and explicit measures.

For descriptive statistics of men's and women's implicit and explicit gender self-concepts, see Table 1. Both implicit and explicit measures follow a similar pattern among men and women. Across the whole sample, we ran Spearman's correlations between the two implicit measures (self-male and self-female association) and the two explicit measures (identification with men and with women). We found, as one might expect, that the self-male association correlated positively with male gender identification (rho = .46, $p < .001$) and negatively with female gender identification (rho = -.47, $p < .001$). The self-female association, on the other hand, correlated positively with female gender identification (rho = .40, $p < .001$) and negatively with male gender identification (rho = -.36, $p < .001$). All correlations were moderate, in

line with previous literature regarding the correlation between explicit and implicit measures (Payne et al., 2008), suggesting that implicit associations about one's own gender somewhat converge with explicit gender identification measures, but that explicit and implicit measures did tap into different thought processes.

Table 1

Means and Standard Deviations of Implicit and Explicit Self-Concepts in Study 1, Split by Men and Women

Measure	Men (*M* (*SD*))	Women (*M* (*SD*))
Implicit female self-concept	-.20 (0.65)	.99 (1.01)
Implicit male self-concept	.93 (1.21)	-.49 (0.82)
Explicit identification with women	2.07 (1.00)	6.22 (0.70)
Explicit identification with men	6.11 (0.93)	2.22 (1.11)

Multiple identification approach to own implicit gender

Given that we had found a significant relationship between the implicit and explicit measures of gender identification, we concluded that we could explore whether we could replicate the findings of Wickham et al. (2022) using an implicit, rather than explicit measure. First, we examined whether the correlation between implicit gender self-concepts (male and female) would be moderate (and negative), which would suggest that seeing oneself as male or as female are related but not polar opposites in people's implicit associations. Indeed, a Pearson's correlation revealed a significant, moderate, and negative correlation ($r = -.38$, $p < .001$), replicating previous findings using an explicit measure of gender identification.

Next, we tested a multiple identification model of gender (Wickham et al., 2022) using the implicit measures, to see whether we could replicate the finding that a binary model of gender is not the best

fit for people's (implicit) gender self-concept. As such, we ran a series of two-step cluster analyses with different numbers of clusters (two, three, four, and five). We found that a three-cluster model was the best fit in terms of BIC and silhouette of cohesion and separation combined (*BIC* = 135.09; .26 ≤ silhouette ≤ .50). Our results therefore suggest that a two-cluster (i.e., binary) model was not the best fit for our implicit data, replicating previous findings using explicit data.

Cluster centres and percentages of participants in each cluster can be found in Table 2. We interpreted d scores that were relatively close to zero to represent medium identification with a gender group, given that a *d* score close to zero represents relatively equal ease of learning both the congruent and incongruent rules. On the other hand, we interpreted d scores larger than 0.5 as high identification with a gender group, and lower than -0.5 as low identification with a gender group. We thus identified the three clusters to represent: 1) high female, low male identifiers, 2) medium female, medium male identifiers, and 3) medium female, high male identifiers. This is consistent with a study by Wickham et al. (2022) which also found a three-category model of explicit gender identification when their sample had an underrepresentation of men, which was also the case in this study. Our cluster centres also suggest that our sample's implicit female self-concept was relatively high, likely because of our sample consisting of mostly women. This further corroborates the idea that a study with a more gender-equal sample may have shown results that were more consistent with previous findings.

Table 2

Cluster Centres of Implicit Self-Concepts and Percentages of Overall Sample in Each Cluster, Study 1

	Cluster 1 (high female, low male)	Cluster 2 (medium female, medium male)	Cluster 3 (medium female, high male)
Cluster centre of implicit female self-concept	1.99	0.24	-0.07
Cluster centre of implicit male self-concept	-0.79	-0.30	1.64
Percentage of overall sample	27.88%	55.77%	16.35%

Implicit associations of other people's genders

Next, we explored participants' implicit associations about other people's (men's, women's) genders. Descriptive statistics (see Table 3) showed that participants generally highly associated men with maleness and women with femaleness. Furthermore, participants generally did not associate men with femaleness and women with maleness. However, we suspected that correlations between men-male and men-female, as well as women-female and women-male associations would be moderately negative at most, just as for implicit associations of one's own gender. Indeed, Pearson's correlations revealed a significant, weak and negative correlation between men-male and men-female associations (r = -.26, p = .007), as well as a moderate and negative correlation between women-female and women-male associations (r = -.35, p = .004). This suggests that, while women are seen as more female than male (and men are seen as more

male than female), the concepts of "male" and "female" are not polar opposites in terms of people's implicit associations. Thus, participants' concepts of men's and women's genders were generally not strictly binary, which we suspected. This finding suggests that implicit concepts of "female" and "male" are not implicitly seen as complete opposites. In other words, participants may be able to conceive of people who are female but also a little male, suggesting a less strictly binary worldview than one might assume.

Table 3

Means and Standard Deviations of Strength of Associations of Others' Genders, Study 1

Implicit association	Mean *d* score (*SD*)
Men are male	0.45 (1.21)
Men are female	-0.26 (1.02)
Women are female	0.96 (1.24)
Women are male	-0.49 (0.87)

Men's and women's views of ingroup and outgroup gender

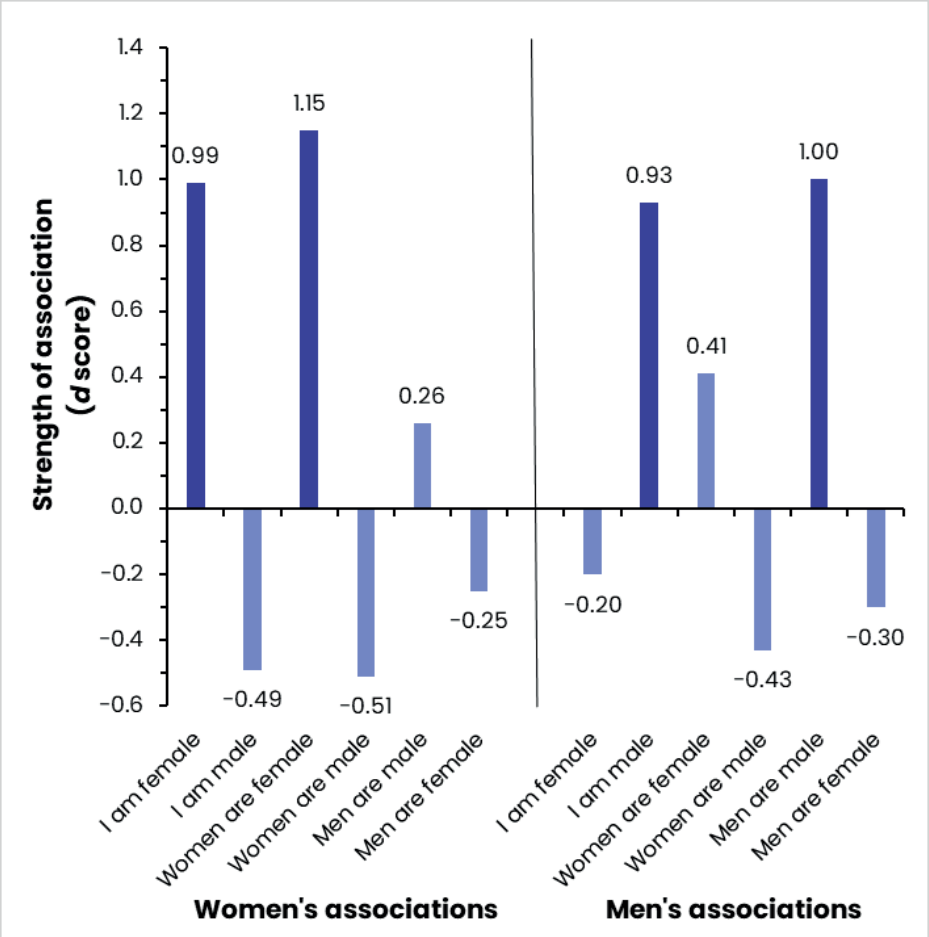
Lastly, we wanted to explore whether views about other people's gender were dependent on the participant's own group membership, that is, whether there were differences between men and women. As such, we tested whether men's and women's implicit associations of other people's gender differed by means of a one-way ANOVA (4 DVs: men-male association, men-female association, women-female association, women-male association). We found a significant difference between men and women in the men-male association ($F(1, 102) = 7.88$, $p = .006$, $\eta_p^2 = 0.07$), such that men ($M = 1.00$, $SD = 1.21$) tended to associate men with maleness more than women ($M = 0.26$, $SD = 1.16$). Similarly, we found a significant difference

between men and women in the women-female association ($F(1, 102) = 7.40$, $p = .008$, $\eta_p^2 = 0.07$), such that women ($M = 1.15$, $SD = 1.21$) tended to associate women with femaleness more than men ($M = 0.41$, $SD = 1.16$). We did not find any differences between groups in the other two associations ($ps \geq .660$). This suggests that there was an ingroup effect, whereby women have stronger associations about women being female than men, and men have stronger associations about men being male than women.

Given that we had found this ingroup effect, we further compared all implicit associations (regarding the self and others) between groups. The pattern of results (see Figure 2) suggested that women's associations about themselves being female and women being female, and men's associations about themselves being male and men being male were very strong ($ds \geq 0.93$). On the other hand, all other associations were comparatively weak ($ds \leq \pm 0.51$). Indeed, follow-up paired samples t-tests revealed no significant differences between women's associations about themselves and other women being female ($p = .319$), while there were significant differences with all other associations (all $ps < .001$). Similarly, men's associations about themselves and other men being male did not significantly differ ($p = .820$), while all other associations did (all $ps \leq .042$). This shows that participants implicitly view themselves similarly to how they view their ingroup.[6]

Figure 2

Strengths of All Associations in Study 1, Split by Men and Women



Note. The darker bars represent the strongest associations for each gender group (men and women). These associations did not differ significantly from each other but differed significantly from all other associations (lighter colour).

Together, these findings suggest an ingroup effect, whereby participants' view of their own gender is highly related to how they view their ingroup, but less related to how they view their outgroup. According to theory (Simon, 1992), one might expect that people would

see their outgroup as more homogenous than their ingroup, and very much distinct from the ingroup. Had that been the case, we would have found that men had strong associations of women being female and not male (indicated by a large positive women-female *d* score and a large negative women-male *d* score). Comparatively, we would have found that men had weaker associations about men being male and not female (and vice versa for women's associations). However, what we found was the opposite, indicating instead that participants implicitly viewed their ingroup as more homogenous (large positive women-female d score for women, and large positive men-male d score for men) than their outgroup. Group distinctiveness was also lower than one might expect based on theory (relatively small negative men-female *d* score for women, and relatively small negative women-male *d* score for men). Our results may thus provide a different insight into self-categorisation and group formation (Turner & Reynolds, 2011) when it comes to the social category of gender. In Study 2, our aim was to replicate results from Study 1 using a longer and more controlled experiment.

Study 2

Methods

Participants

We recruited 99 participants among Bachelor students of the Psychology course at Utrecht University, and among friends of those students, but had to exclude five participants based on low quality data. Our final sample size was thus $N = 94$ of which 78 were self-identified women and 16 were self-identified men[7]. Due to an error during set-up, participants' age and other demographics are unknown.

Materials

Implicit gender associations. Participants completed the same

IRAP as in Study 1. However, we included a larger number of stimuli (see Appendix B) and therefore each block included 144 trials.

Explicit gender identification. To measure participants' level of gender identification with each binary gender group (men, women), we included 12 items from Wickham et al. (2022), six about men and six about women. Items were measured on a 7-point Likert scale (1 = "completely disagree", 7 = "completely agree"). Examples included "I identify with [wo] men" and "[Wo]men are an important reflection of who I am"[8] (items about men, α = .94; items about women, α = .93).[9]

Procedure

We recruited participants via the student participation system of Utrecht University, and among the student experimenters' own networks. Participants were redirected to the experiment on Gorilla Experiment Builder (Anwyl-Irvine et al., 2020), and gave informed consent. All participants completed the experiment in Dutch. They were asked to indicate their gender, after which they did the implicit gender associations task. Next, they were asked to answer the self-report questions, in the order in which they appear in this chapter. They were asked to indicate whether they are colourblind, given that one pair of stimuli was the same for men and women, except for the colour (pink and blue). Lastly, they were debriefed and, if they wished, received compensation in the form of study credit.

Results and Discussion

Implicit associations of own gender

Following Study 1, we investigated whether implicit (*d* scores) and explicit gender self-concept measures converged with one another but remained distinct. For descriptive statistics of men's and women's implicit and explicit gender self-concepts, see Table 4. Replicating results from Study 1, Pearson's correlations showed a significant positive correlation

of self-male association with explicit male gender identification (r = .38, p < .001), as well as a negative correlation with explicit female gender identification (r = -.37, p < .001). Similarly, the self-female association correlated positively with explicit female gender identification (r = .28, p = .006) and negatively with explicit male gender identification (r = -.25, p = .014). These findings suggest that the implicit and explicit measures capture somewhat similar but distinct constructs.

Table 4

Means and Standard Deviations of Implicit and Explicit Self-Concepts in Study 2, Split by Men and Women

Measure	Men (*M* (*SD*))	Women (*M* (*SD*))
Implicit female self-concept	-.02 (0.53)	.62 (0.58)
Implicit male self-concept	.53 (0.62)	-.13 (0.50)
Explicit identification with women	2.18 (0.85)	5.79 (0.81)
Explicit identification with men	5.45 (1.06)	1.76 (0.63)

Multiple identification approach to own implicit gender

Following Study 1, and the methods of Wickham et al. (2022; using an explicit measure) we tested whether implicit gender self-concept was binary. First, a Pearson's correlation revealed a significant, moderate-high, and negative correlation between implicit female self-concept and implicit male self-concept (r = -.52, p < .001). This suggests that seeing oneself as male or as female are highly related but not polar opposites in people's implicit associations.

Replicating findings from Study 1, a series of two-step cluster analyses with different numbers of clusters revealed that a three-cluster model was the best fit in terms of *BIC* and silhouette of cohesion and separation combined (BIC = 113.93, silhouette ≥ .50). Thus, implicit gender self-concepts were not binary (i.e., a two-cluster model was not the best fit).

Cluster centres and percentages of participants in each cluster can be found in Table 5. We interpreted d scores in the same way as in Study 1. We identified the three clusters to represent: 1) high female, low male identifiers, 2) high female, medium male identifiers, and 3) medium female, high male identifiers. These resemble the clusters found in Study 1, even though *d* scores in Study 2 were overall lower than in Study 1. We attribute the d scores being lower to participants being able to learn the rules of the reaction time task more easily because they were in a quieter, more controlled environment and were performing a longer task with more trials (van Nunspeet et al., 2014). This likely affected their accuracy such that they were more accurate in incongruent trials, where the rule is more difficult to learn.

Table 5

Cluster Centres of Implicit Self-Concepts and Percentages of Overall Sample in Each Cluster

	Cluster 1 (high female, low male)	Cluster 2 (high female, medium male)	Cluster 3 (medium female, high male)
Cluster centre of implicit female s elf-concept	1.10	0.60	-0.20
Cluster centre of implicit male self-concept	-0.71	0.01	0.57
Percentage of overall sample	24.47%	48.94%	26.60%

Implicit associations of other people's genders

We explored whether participants' implicit associations about other people's genders (for descriptive statistics, see Table 6) were similar in Study 2 as in Study 1. We indeed found that participants generally associated men with maleness and women with femaleness. Furthermore, participants generally did not associate men with femaleness and women with maleness, replicating findings from Study 1. A Pearson's correlation revealed a significant, moderate, and negative correlation between women-female and women-male associations (r = -.42, p < .001), and a non-significant, weak and negative correlation between men-male and men-female associations (r = -.18, p = .077). Our findings in Study 2, much like findings of Study 1, suggest that maleness and femaleness are not seen as polar opposites in people's implicit associations. However, the correlation between men-male and men-female associations being non-significant differs from Study 1. This is likely due to small methodological differences between studies, namely a combination of the underrepresentation of men in our sample and the longer task which was related to lower d scores in Study 2 as compared to Study 1. Specifically, we had previously found that men tended to have stronger men-male associations than women; thus, having fewer men in our sample meant that the overall men-male association was relatively low. This, coupled with the lower d scores in general, likely meant that we could not find our previously found effect for the men-male association. We suspect that we would have found a moderate negative correlation between men-male and men-female associations in a more gender-equal sample, which would have replicated results from Study 1.

Table 6

Means and Standard Deviations of Strength of Associations of Others' Genders

Implicit association	Mean *d* score (*SD*)
Men are male	0.23 (1.21)
Men are female	-0.32 (1.02)
Women are female	0.53 (1.24)
Women are male	-0.15 (0.87)

Men's and women's views of ingroup and outgroup gender

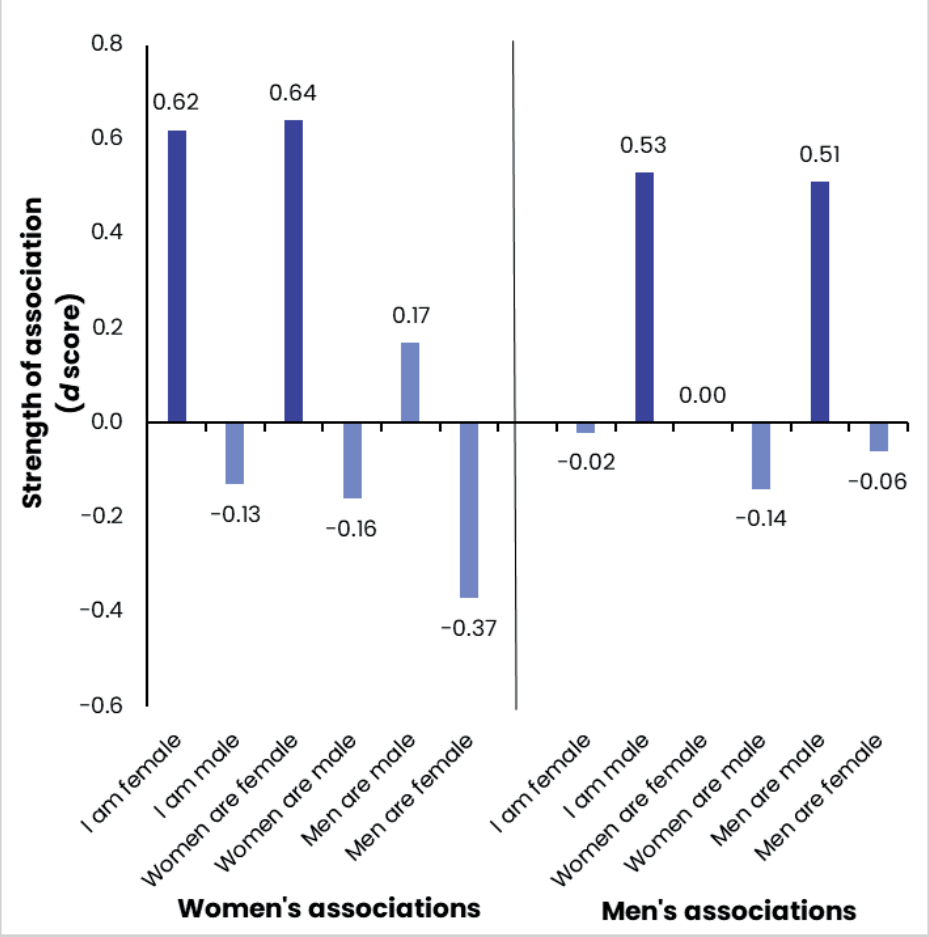
Lastly, we investigated whether views about other people's gender were dependent on the participant's own ingroup, that is, whether there were differences between men and women, just as in Study 1. A one-way ANOVA revealed significant differences between men and women in their men-male ($F(1, 92) = 5.72$, $p = .019$, $\eta_p^2 = 0.06$), men-female ($F(1, 92) = 5.30$, $p = .024$, $\eta_p^2 = 0.05$), and women-female associations ($F(1, 92) = 14.16$, $p < .001$, $\eta_p^2 = 0.13$), but not their women-male associations ($p = .942$). Specifically, women tended to have stronger women-female associations ($M = 0.64$, $SD = 0.61$) than men ($M = 0.00$, $SD = 0.66$), weaker men-male associations ($M = 0.17$, $SD = 0.52$) than men ($M = 0.51$, $SD = 0.50$), and stronger men-female associations ($M = -0.37$, $SD = 0.49$) than men ($M = -0.06$, $SD = 0.50$). This shows the same ingroup effect as in Study 1, whereby women have stronger associations about women being female than men, and men have stronger associations about men being male than women. It also shows an additional effect, whereby women have a stronger association about men not being female than men, though this effect is relatively weak.

We then compared all implicit associations (regarding the self and others) between groups. The pattern of results (see Figure 3) suggested that women's associations about themselves being

female and women being female, as well as men's associations about themselves being male and men being male were very strong ($d \geq 0.51$). All other associations were comparatively weak ($ds \leq \pm 0.37$). Follow-up paired samples t-tests revealed no significant differences between women's associations about themselves and other women being female ($p = .658$), while all other associations significantly differed (all $ps \leq .001$). Similarly, men's associations about themselves and other men being male did not differ ($p = .773$), while all other associations did (all $ps \leq .001$). This shows that participants implicitly view themselves similarly to how they view their ingroup, following the findings of Study 1.[10]

Figure 3

Strengths of All Associations in Study 2, Split by Men and Women



Note. The darker bars represent the strongest associations for each gender group (men and women). These associations did not differ significantly from each other but differed significantly from all other associations (in lighter colour).

General Discussion

Across two empirical studies, we shed a light on implicit gender self-concepts and other-concepts. The aims across both studies were to replicate previous findings that gender identification is not binary

by using an implicit measure, and to add to previous literature about gender self-concepts by investigating implicit concepts of others' genders. We found consistent results across both studies: that a binary model of implicit gender self-concepts is not the best fit for our data, that concepts of "male" and "female" are not implicitly seen as polar opposites, and that implicit gender self-concepts are related to implicit concepts of the ingroup's, but less so the outgroup's, gender. These findings add to the existing literature on gender non-conformity, social cognition and self-categorisation.

Implicit gender self-concepts

Previous literature (Abed et al., 2019; Andrews et al., 2019; Baiocco et al., 2022; Endendijk et al., 2019; Joel et al., 2014) has consistently shown that gender identification is not binary, in that a substantial proportion of the population can be seen as "gender non-conforming": they identify with both men or women, with neither, and/or with another gender altogether. However, these studies have used explicit measures, such as questionnaires, to come to these conclusions. We added to this literature by showing that implicit gender self-concepts, that is, implicit associations of the self with men or women, are similarly not binary. Specifically, we found that a binary model of implicit gender was not the best fit for our data, and that around half of participants in both studies seemed to implicitly see themselves as both somewhat male and somewhat female. Given that implicit and explicit measures seem to capture distinct but related constructs (Payne et al., 2008; and we also showed this in our own study), this is an interesting finding. It suggests that people not only explicitly see themselves in a gender non-conforming way, they have also internalised this implicitly, which may provide insight into their automatic, unconscious or "bottom-up" processing of gender (Gawronski & Payne, 2010). This is further evidence that the societal debate about the gender binary (Cooper, 2019) and

the increasing number of gender-inclusive initiatives (Hord, 2016; Pun et al., 2017; Reis et al., 2018) reflect that people do not see themselves in a strictly binary way.

Male is not the opposite of female

Previous work by Wickham et al. (2022) showed that gender identification with men and gender identification with women were not polar opposites of one another, in that they were only weakly or moderately (and negatively) related. Using our implicit measure, we showed that this was also the case for implicit male and female self-concepts, as well as concepts of others being male and female. Specifically, across both studies we showed that implicit associations of the self, women in general, and men in general, with "male" and "female" were moderately, negatively correlated. In a strictly binary world where men and women are seen as polar opposite groups, you would expect to find a (very) large negative correlation between the two. Thus, a moderate correlation suggests that the concepts of "male" and "female" are not polar opposites, and that participants could implicitly conceive of themselves, as well as others, as both somewhat male and somewhat female. This has implications for our understanding of social cognition regarding gender (Bussey & Bandura, 1999), as it shows that gender differentiation (i.e., the understanding that a person can be one of two genders: male or female) may not be as clear-cut as previously thought.

The relationship between the self, the ingroup and the outgroup

Self-categorisation theory (Turner & Reynolds, 2011) tells us that we are more likely to categorise ourselves as members of a social group when there is a high level of fit with the ingroup and a high level of distinctiveness from the outgroup. As such, our social self-concepts are tied to our concepts of the ingroup and the outgroup. We thus expected that implicit gender self-concepts would be related to implicit associations

about both the ingroup's and outgroup's genders. Instead, we found that implicit gender self-concept was more related to implicit gender concepts about the ingroup, than about the outgroup. Furthermore, it has been shown that people tend to view the outgroup as more homogenous than the ingroup (Simon, 1992). Based on this, one would expect that implicit associations about the outgroup's gender would be stronger than implicit associations about the ingroup, since the ingroup may be seen as more heterogeneous. Instead, we found that implicit gender associations about the ingroup were stronger than those about the outgroup.

Specifically, across both studies, we found that women tended to have the strongest associations about themselves being female and women in general being female, while all other associations were comparatively weak. Similarly, men tended to have the strongest associations about themselves being male and men in general being male, with all other associations being comparatively weak. Our findings thus suggest that implicitly viewing yourself as a member of your gender group is, at baseline, most linked to the prototypicality of your gender ingroup, and somewhat less the distinctiveness from your outgroup. This may differ under different conditions. For instance, if group distinctiveness had been threatened, participants may have seen the outgroup as more homogeneous and distinct from themselves and their ingroup (Spears et al., 1997). In Study 2, we arguably found a weak effect of group distinctiveness, whereby women had a stronger association about men not being female than men. A potential avenue for future research is to further investigate whether this effect is present under various different conditions.

Suggestions for future research

In this chapter we took a first step towards researching whether our views of other people's genders follow a strictly binary view of gender. Our findings, that "male" and "female" are not implicitly seen

as polar opposites, and that our implicit associations of others' genders are dependent on our own viewpoint (i.e., who is the ingroup and who is the outgroup), suggests that we do not see others in a strictly binary way. The *gender binary* is the view that humans all fit into one of two categories, men and women, that these categories are biologically meaningful, and that this view is natural and unmalleable (Hyde et al., 2019). Given that gender categorisation is usually studied in a binary way (e.g., Kecskés-Kovács et al., 2013; Mouchetant-Rostaing et al., 2000), we think that the findings in this chapter indicate that more research into gender categorisation *beyond the binary* is warranted. Gaining a better understanding of the conditions under which people categorise others in a binary way, and whether this is dependent on the type of information provided or the social context, would give us further insight into people's understanding of gender.

Limitations

There were several limitations that our studies share. First, in both of our studies, we were lacking demographic information about age, and other identities that may intersect with gender (e.g., race, sexuality, disability) and may thus affect participants' understanding of gender (Cooper, 2016).

Second, while it is a strength that each of our studies were conducted in different environments with different samples, each of our studies also had certain limitations regarding where data was collected and among whom. Study 1 was conducted at a popular science festival, which meant we were able to collect data among a highly motivated sample (as evident from conversations with participants by the first author) within the space of a few hours. However, the festival was far from a controlled environment, given noise levels and large volume of people walking past participants. Furthermore, we had to keep participation time short due to festival guidelines, meaning our reaction

time task contained a relatively small number of trials. Study 2, on the other hand, was conducted among a sample of mostly Psychology students (a highly homogeneous sample) who were able to participate from the comfort of their own homes, and could complete a longer version of the task. We think that, due to these discrepancies in data collection, we saw differences in results of the studies. Specifically, in Study 1, implicit associations were stronger (larger *d* scores) than in Study 2. We think this is because it was harder to do the task well in the hustle and bustle of the festival, and this was likely especially the case for incongruent trials where the rule (e.g. "women are not female") was harder to learn than in congruent trials. Nonetheless, even with these discrepancies, we largely replicated results across both studies.

Third, across both studies, we had an underrepresentation of men in our samples. As shown in previous research (Wickham et al., 2022), an underrepresentation of one gender group can skew results regarding gender identification. Specifically, across both studies we found a three-cluster model of implicit gender self-concept, rather than a four-cluster model which one would expect based on some previous literature (e.g., Andrews et al., 2019). However, our findings were consistent with a study by Wickham et al. (2022) which also found a three-category model of explicit gender identification when their sample had an underrepresentation of men. Furthermore, we found slightly different results in terms of the interpretations of each cluster. For instance, in Study 2, there was no group that had a very low implicit female self-concept, while in Study 1, there was. We think this is in part an artifact because of the underrepresentation of men in our samples, and partly due to differences in d scores between studies affecting how we interpreted our results. Nonetheless, we achieved our goal of replicating, across both studies, the finding that implicit gender self-concept was not binary (i.e., a two-cluster model was not the best fit).

Conclusion

Across two empirical studies, we added to existing literature about gender non-conformity, gender social cognition, and gender self-categorisation, by exploring people's implicit gender self-concepts and implicit concepts of others' genders. Using a novel measure, we found further evidence that people do not self-categorise in a strictly binary way, as has been shown in previous literature. Additionally, we show that people's cognitive understanding of gender is not strictly binary, in that the concepts of "male" and "female" are not implicitly seen as polar opposites of one another. Lastly, we found that people's implicit views of other people's genders were dependent on their own gender group. We conclude that researchers should further investigate how we see others' genders, or how we categorise others by gender, beyond the current binary framework.

Appendix A: Stimuli used in IRAP, Study 1

Words	Female	Male
	Zij (She)	Hij (He)
	Mevr. (Mrs.)	Dhr. (Mr.)
Icons	Female	Male




Appendix B: Additional stimuli used in IRAP, Study 2

Words	Female	Male
	Meisje (girl)	Jongen (boy)
	Vrouwelijk (female)	Mannelijk (male)
	Zus (sister)	Broer (brother)
	Moeder (mother)	Vader (father)
	Nicht (niece)	Neef (nephew)
	Tante (aunt)	Oom (uncle)
	Actrice (actress)	Acteur (actor)
	Prinses (princess)	Prins (prince)
	Koningin (Queen)	Koning (King)
	Stewardess	Steward

Icons	Female	Male
		
		
		
		












Footnotes

[1] We asked them to answer the question (in Dutch) "Wat is jouw gender?" with the answer options "Man", "Vrouw" or "Anders".

[2] We also ran our analyses excluding participants who had consumed more than three units of alcohol, but found no differences in results. We thus decided to include all participants, regardless of units of alcohol consumed.

[3] In Dutch, "zij" and "hij".

[4] In Dutch, "mannen =", "vrouwen =", "ik =".

[5] In Dutch: "Ik identificeer me met mannen" and "Ik identificeer me

met vrouwen"; Likert scale going from 1 ("helemaal mee oneens") to 7 ("helemaal mee eens").

[6] We also compared all other effects with each other. These results are beyond the scope of this chapter but can be found in the Supporting Information.

[7] We asked them to answer the question (in Dutch) "Wat is jouw gender?" with the answer options "Man", "Vrouw" or "Anders".

[8] In Dutch, "Ik identificeer mij met mannen/vrouwen" and "Mannen/vrouwen zijn een belangrijke representatie van wie ik ben"

[9] We also asked participants questions about their beliefs regarding gender essentialism, precarious manhood, precarious womanhood, and prescriptive and descriptive stereotypes about men and women. The findings from these measures are beyond the scope of this chapter, but can be found in the Supporting Information.

[10] We also compared all other effects with each other. These results are beyond the scope of this chapter but can be found in the Supporting Information.

Chapter 6: Gender Categorisation Based on Biological and Social Information

Beyond the Gender Binary

"Plucked her eyebrows on the way, shaved her legs and then he was a she"

Take a Walk on the Wild Side – Lou Reed

Author contributions: Miriam I. Wickham (Conceptualisation; Data curation; Formal analysis; Investigation; Methodology; Validation; Writing – original draft; Writing – review & editing); Félice van Nunspeet (Conceptualisation; Methodology; Validation; Writing – review & editing); Naomi Ellemers (Conceptualisation; Funding acquisition; Methodology; Validation; Writing – review & editing)

Abstract

We are witnessing a polarising gender debate in society right now, whereby one side is arguing for the biological basis of gender, and the other for the social construction of gender. Across four studies, we set out to investigate people's *explicit* and *implicit* ideas about gender and whether these may be more nuanced than the public debate makes it seem. Specifically, we asked participants to reflect and report (Studies 1a and 1b; explicit measure) on the kinds of decisions they make during gender categorisation based on biological (e.g., facial structure) and social (e.g., clothing) information. Participants reported that both biological and social information was important when they categorised someone by gender. We also asked participants to perform a gender categorisation task (Studies 2a and 2b; implicit measure) where they were presented with a biological stimulus (a face) and a social stimulus (a shoe), and we investigated both their behaviour and attention (eye-tracking). We found that participants based their gender categorisations on both the social and biological stimuli, and paid attention to both. We also found that their behaviour and attention towards the social stimulus was higher in certain social contexts, showing that participants were also (implicitly) aware of the importance of social information for expressing identity. Overall, our findings suggest that people actually consider gender as partly biologically based and partly socially constructed, showing a nuanced (explicit and implicit) view of gender conceptualization. These findings can inform the polarised societal debate about gender.

In a 1973 interview, Russell Harty asked David Bowie, who was wearing colourful high-heeled shoes: "And how about the shoes? Are those men's shoes, or women's shoes, or bisexual shoes?" To this, David Bowie replied: "They're shoes-shoes, silly!" (Bailey et al., 1972-1981). Fifty years later, we still see clothes as a form of gender expression which is mostly binary (clothes are either for men or for women). However, there are still those who, like David Bowie, step out of gender norms in terms of the clothes they wear. In fact, in recent years, consumers' interest in "genderless fashion" has been found to have increased dramatically (Kim et al., 2022), highlighting society's current shift away from a strictly binary view of gender and towards a more inclusive understanding of both gender identity and expression.

At the same time, there is a large public debate about the conceptualization of gender in society (Cooper, 2019), demonstrating that the shift away from the gender binary is not a smooth one. Specifically, one side of the debate argues that gender is entirely biologically based, and therefore that gender is a rigid, binary category. On the other hand, the opposing side argues that gender is socially constructed, that the biological base is not decisive, and thus that non-binary gender options are feasible too. The assumption underlying this debate seems to be that most people consider *either* biological *or* social information to be decisive in determining someone's gender, while the reality is likely more nuanced. To investigate people's gender conceptualizations, we took a step back from the societal debate of how gender *should* be conceptualized, and investigated how people *currently* use biological or social information to understand other people's genders. Specifically, we first asked participants to reflect on the kinds of information they use (biological and social) when categorising others by gender. Subsequently, we asked participants to perform a gender categorisation task based on both biological and

social cues. By doing so, we were able to explore people's (explicit and implicit) conceptualizations of gender, and whether they use biological or social cues to categorise others, or a little of both. We anticipated that people's concepts of gender would be more nuanced than the public debate would have us believe, such that they make use of both biological *and* social information to understand someone's gender.

The public debate about non-binary gender

Gender as a social category is highly morally charged, whereby there are strict gender norms and those who break them are discriminated against. For instance, women who have higher societal status than their male partners are penalized (Vink et al., 2023), and high-status women often have to distance themselves from their gender group for continued career success (Faniko et al., 2017). Men face backlash for being modest rather than assertive (Moss-Racusin et al., 2010), and feel they cannot ask for psychological help in times of distress without compromising their masculinity (Keohane & Richardson, 2018). People who identify in a gender non-conforming way feel less included by society (Wickham et al., 2023) and transgender people deal with harassment, violence and economic discrimination (Lombardi, 2009). All these findings indicate that breaking gender norms has negative consequences for people of all gender identities, which is why there are calls to change our current gender system.

Our current system is a gender binary system, though this is in the process of shifting. The *gender binary* is the view that humans all fit into one of two categories, men and women, that these categories are biologically meaningful, and that this view is natural and unmalleable (Hyde et al., 2019). This is still the pervasive view; for instance, in most forms that we fill out, we are asked to indicate our binary gender and are often not given any other gender options (Morgenroth & Ryan, 2018). However, in recent years, a public debate about our current gender

system has emerged, with two opposing sides (Cooper, 2019). One side invokes arguments about gender being based on biological (binary) sex to argue that we must keep the rigid binary system, and the other argues that gender is an unfixed, diverse identity, and thus that the current system needs to change. For instance, we have seen a societal increase in initiatives that aim to include people with diverse gender identities (Pun et al., 2017), as well as these kinds of initiatives being shot down by those who believe gender to be entirely biologically based (Blumell et al., 2019). This public debate would make it seem like most people use *either* biological *or* social information to categorise others by gender, while people most likely use both. In this chapter, we sought to investigate that nuance by exploring how people use biological *and* social information to understand other people's genders.

We were also interested in both people's explicit estimates of how they use biological and social information to categorise others by gender, as well as a more implicit measure: their actual behaviour of which they may not be aware. This is because we considered that there could be a discrepancy between participants' explicit estimations and their actual behaviour, particularly since the topic of investigation could be seen as polarising. Specifically, questions about how individuals categorise others by gender may invite answers based on participants' ideology and thus what they perceive to be socially desirable (Paulhus, 2017). Furthermore, individuals are sometimes not aware of their own underlying thought processes and are thus unable to estimate them correctly (Nisbett & Wilson, 1977; Rule & Sutherland, 2017). Investigating whether findings from explicit and implicit measures are similar or not can give us a more complete picture of what people do and what they *think* they do. As such, in this research, we explored gender categorisation based on biological and social information, both explicitly (through self-report) as well as more implicitly (through a gender categorisation task).

Gender categorisation from a non-binary perspective

Gender categorisation is the process by which we distinguish other people's genders. It is a highly pervasive form of social categorisation, in that categorising someone by gender often occurs implicitly, even when this categorisation is not relevant to the situation (e.g., Ito & Urland, 2003). Furthermore, gender categorisation is at the core of gender stereotypes, prejudice and intergroup conflict (Krueger & DiDonato, 2008), making it a relevant process to study in order to better understand changing societal norms around gender.

However, though societal norms around gender seem to be shifting, research into gender categorisation is still mostly being done from the gender binary perspective, whereby there are only two genders, and these are based solely on biological sex. For instance, most research into gender categorisation operates from the assumption that we (only) use biological information to categorise others by gender (e.g., Freeman et al., 2010; Kecskés-Kovács et al., 2013; Latinus et al., 2010; Mouchetant-Rostaing et al., 2000), such that experiments include only biological features (e.g., faces devoid of make-up or other social markers) as stimuli. Research into the importance of social information in gender categorisation has been limited (e.g., Huart et al., 2005; Leinbach & Fagot, 1993; Martin & Halverson, 1983; Weißflog & Grigoryan, 2024), but has shown that social information can affect or sometimes override biological information during gender categorisation. Thus, including both biological and social information in gender categorisation research can provide us with a more complete understanding of how we view others' genders.

In our research, besides including both biological and social features in our research, we also investigated the influence of social context. We did this because we considered that gender categorisation does not occur in a vacuum, and people may be affected by what they

perceive to be important based on the social context around them. While research into the effects of social context on categorisation has been limited, social context has been found to affect the salience of social categories. Specifically, a typically salient social category (race) is seen by participants as less important when the social context dictates that another category is more important (Van Bavel & Cunningham, 2010). In our latter studies, we thus introduced different social contexts in which binary gender was implied to be more or less important and investigated how these affected gender categorisations.

The present studies

Across four studies, we investigated how biological cues of sex and social cues of gender are used and interact when people categorise others by gender. In the first two studies (1a and 1b) we focused on people's explicit estimates of how they categorise others by gender. Specifically, participants were asked to rate biological and social cues of gender by how important they thought they were when categorising others by gender, and we compared which cues they considered to be most important. We anticipated that participants would estimate biological cues of sex to be more important than social cues of gender, given the pervasive societal norm about gender still being binary and biologically essentialised. In the latter two studies (2a and 2b) we first focused on participants' actual behaviour during a gender categorisation task where we combined biological cues (faces) with social cues (shoes). Then we asked participants to estimate to what extent they used the biological or social cue during the gender categorisation task. We were thus able to compare their actual behaviour (more implicit measure) to their self-reflection on the way they categorise individuals (explicit measure). In the final study (2b), we also compared participants' attention to different stimuli (measured with eye-tracking) to their behaviour, to get more insight into the underlying processes of their gender categorisation.

Additionally, in Studies 2a and 2b, we researched how social context may affect participants' gender categorisation, to see under which conditions social cues are most likely to override biological cues. Specifically, we investigated how participants' categorisations differed in different contexts, where (binary) gender categorisations were portrayed as more or less important.

Study 1a

Method

Participants

Based on a power analysis, our final sample (N = 84) consisted of a majority of female (71.43%), heterosexual (71.43%), cisgender (94.05%), and White (90.48%) participants. 61.90% of participants were students recruited from KU Leuven, Belgium, while the remaining 38.10% were recruited from Utrecht University, the Netherlands. We found no differences in results between people of different demographics. For more information about our participant sample, including power analysis, for all our studies, see Supplementary Information.

Materials

Measures.

Importance of features. To measure how important participants thought certain features of a person were when categorising others according to gender, we created a list that included both "biological" and "social" cues. We defined biological cues as occurring in the body and which people typically have little influence over besides through plastic surgery (e.g., "thick or thin lips" and "small or large waist"). We based our face-related items on the list of sex differences in facial appearance from Bruce et al. (1993) and based some of our body-related items on the description of sexual dimorphism of

body composition by Wells (2007). Social cues were forms of gender expression that people can easily have influence on (e.g., "presence or absence of make-up", "shoes, heels or no heels"). These were entirely self-written, based on extensive reading of the literature about gender categorisation cues. On a 7-point Likert scale (1 = "Not important at all", 7 = "Very important"), participants rated how important they thought each feature was in deciding someone's gender when they first meet a new person. Because these items were self-constructed, we explored the data from them at item level, rather than constructing scales.

Procedure

We received ethical approval from the ethics board of our local Faculty of Social and Behavioural Sciences, which abides by APA ethical guidelines and the Declaration of Helsinki. Participants completed this study on Qualtrics (https://qualtrics.com/), where they first gave informed consent and were asked to answer all questions without thinking too long about their answers. Next, participants answered questions regarding the importance of different features when categorising others in terms of gender. The items within this survey were randomized. There was an attention check randomly displayed among the different features, namely "Please choose option one". Following this, they answered demographic questions. Lastly, they answered a question with which we checked whether they had participated in another of our studies within the same project[1]. Participants were debriefed and received study credit for their participation.

Results

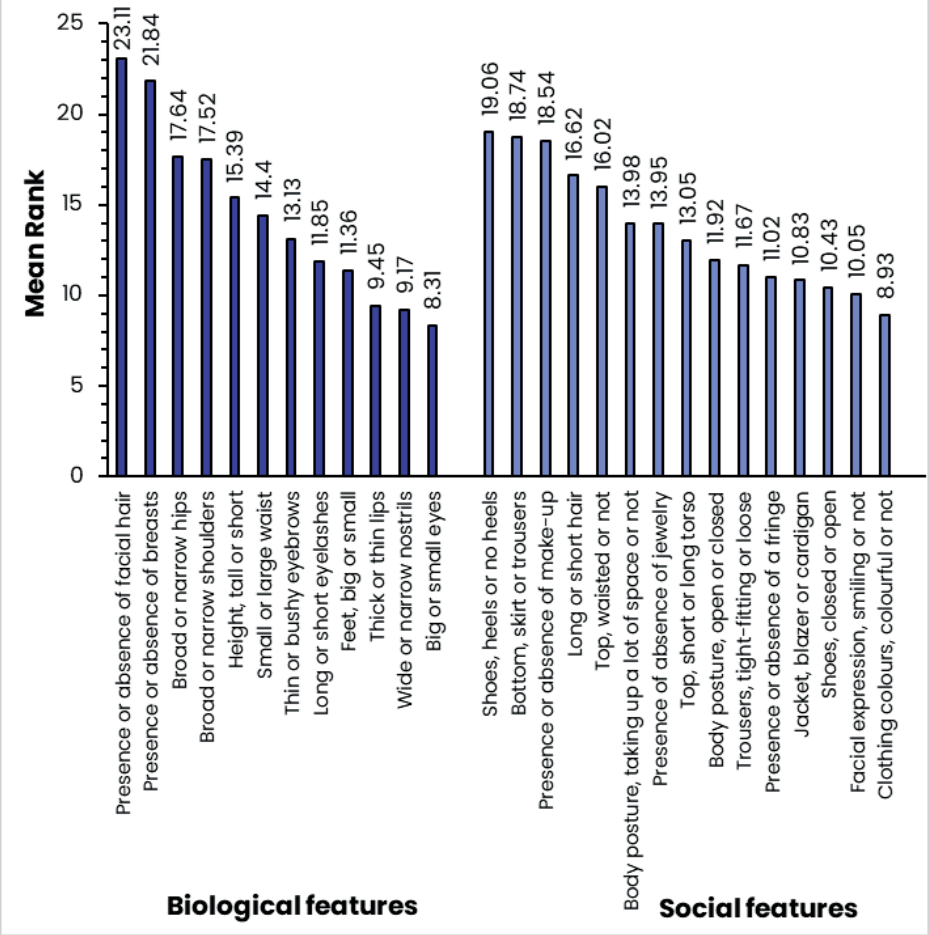
Importance of different features for gender categorisation

We wanted to explore which features were rated as most important for gender categorisation, and whether those features were biological or social. We explored this in two stages. First, we ran

a Friedman test to reveal the ranks of each of the cues (see Figure 1). We found that the cues that were ranked most important ("presence or absence of facial hair" and "presence or absence of breasts") were both biological features, suggesting that biological features may be considered more important for gender categorisation than social features. The highest-ranking social features were not far behind, showing that participants rated both biological and social features as important, and were "shoes, heels or no heels" and "bottom, skirt or trousers". The Friedman test was overall significant ($\chi^2(26) = 636.91$, $p < .001$, $W = 0.29$), showing a main effect of differences between all ranks.

Figure 1

Ranks of Each of the Gendered Features in the Friedmann Test



Second, we ran Wilcoxon signed rank tests to see whether the top ranked biological features differed significantly from the top ranked social features. In a series of four Wilcoxon signed rank tests with Bonferroni correction, we compared each of the top two biological features with each of the top two social features. We found that both of the top two biological features differed significantly from both top two social features in how important participants considered these as cues for gender categorisation. Specifically, facial hair was significantly more

important than heels on shoes (Z = -4.79, p < .001, r = 0.52) and skirts versus trousers (Z = -5.44, p < .001, r = 0.59). Similarly, breasts were significantly more important than heels on shoes (Z = -3.83, p < .001, r = 0.42) and skirts versus trousers (Z = -4.08, p < .001, r = 0.45). This further suggests that biological features are considered more important for gender categorisation than social features.

Discussion

We succeeded in our goals for Study 1a, which were to make an inventory of social and biological cues of gender and explore which features would be rated as most important for gender categorisation by participants.

As expected, we found a large range of importance ratings across both biological and social cues of gender, showing more variation within the biological and social feature categories than between them. This suggests that people view both biological and social features as important in determining someone's gender. However, we also found that the top-rated biological cues of gender (facial hair; breasts) were rated as more important for gender categorisation than the top-rated social cues of gender (heeled shoes or not; skirt or trousers). This suggests that participants view biological cues overall as more important than social cues when reflecting on their own thought processes during gender categorisation.

Given the large range of importance ratings across all features of gender, we wanted to explore how importance ratings of biological and social features would differ if we asked participants to rate them by category (e.g., "biological features of the body"), rather than by specific feature. This is thus what we explored in Study 1b, and we expected that biological features would be rated as more important than social features, overall, but that participants would still rate social cues as important as well. Furthermore, given that beliefs about which

features are most important may be related to ideological beliefs about gender, we also included some additional self-report items in Study 1b. Specifically, we asked participants questions regarding their beliefs about gender being binary and biologically essentialised, and expected these beliefs to correlate with importance ratings of features.

Study 1b

Method

Participants

Our final sample size was *N* = 143. The majority of participants were female (81.82%), heterosexual (69.23%), cisgender (97.20%), White (94.41%), and Psychology students at KU Leuven (73.43%). We found no differences in results between people of different demographics.

Materials

Measures.

Importance of features. To measure how important participants thought certain features of a person were when categorising others according to gender, we asked them to rate, on a 7-point Likert scale (1 = "Not important at all", 7 = "Very important") the following five features: "Biological facial features, such as face shape, shape of the eyes, mouth, chin or jaw", "Accessories around the face, such as make-up, jewelry, or hairstyle", "Biological features of the body, such as height, body shape or cleavage", "Clothes, such as skirts versus trousers, high heels or flat shoes", and "Facial expression or body language, such as how open or closed someone seems". These items are categories that encompass the items from Study 1a.

Binary gender beliefs. To measure whether participants' beliefs about gender were binary (and thus biologically essentialized), we asked participants to rate, on a 7-point Likert scale (1 = "Completely

disagree", 7 = "Completely agree"), how much they agreed with the following statements: "Someone born male is a man, regardless of whether he is wearing a dress", "Someone born female is a woman, regardless of whether she is wearing a suit", and "When I meet someone, I almost always think of them as a man or a woman, and rarely non-binary". We analysed these items separately from each other.[2]

Procedure

The procedure for ethical approval, participant recruitment, online survey platform, and informed consent were the same as in Study 1a. First, participants answered questions regarding the importance of different features when categorising others in terms of gender. Next, participants answered questions about their binary gender beliefs. There was an attention check, namely "Please choose I completely disagree". All items within this survey were randomized. Following this, they answered demographic questions. Lastly, they answered a question with which we checked whether they had participated in another of our studies within the same project.[3] Participants were debriefed and received study credit for their participation.

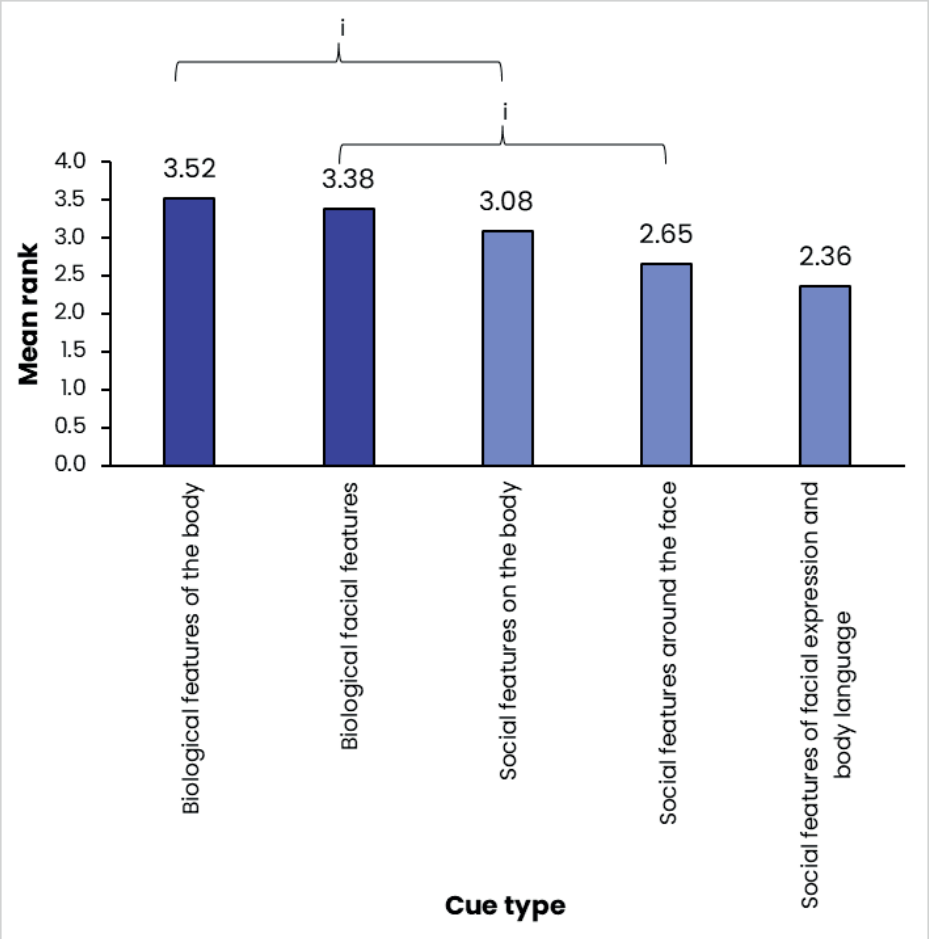
Results

Importance of different features for gender categorisation

Following the data analysis procedure from Study 1a, we analysed this data in two stages. First, a Friedman test revealed the ranks of each of the types of cues (see Figure 2). We found that the cues that were ranked most important were both of the biological features, suggesting that biological features may be considered more important for gender categorisation than social features. The Friedman test was overall significant ($\chi^2(4) = 63.91$, $p < .001$, $W = 0.11$).

Figure 2

Ranks of Each of the Gendered Features in the Friedmann Test



Note. We have indicated which differences we found interesting to further investigate in follow-up tests with the letter "i". We chose these because we wanted to compare the biological and social features of each area: the body and the face.

Second, we ran follow-up Wilcoxon signed rank tests to see whether each of the biological features differed significantly from each of the corresponding social features. In a series of two Wilcoxon signed rank tests with Bonferroni correction, we compared biological features of

the body with social features of the body, as well as biological features of the face with social features of the face. We wanted to see whether, within each location (face, body), there was a significant difference between biological features and social features. We found that biological features of the body were ranked as significantly more important for gender categorisation than social features on the body (Z = -3.20, p = .001, r = 0.26). Similarly, we found that biological features of the face were ranked as significantly more important than social features on and around the face (Z = -4.28, p < .001, r = 0.36). This suggests that, when shown a face or a body, participants would consider biological features more important for gender categorisation than social features. However, as can be seen from the relatively small effect sizes, social features are considered almost as important as biological features, suggesting that people may view gender as both biological as well as social.

Relationship between binary gender beliefs and importance ratings of features

To test whether importance ratings of any of the features correlated with beliefs regarding gender being binary or biologically essentialised, we ran a series of Spearman's correlations (see Table 1). The pattern in the data indicates that importance ratings of biological features tended to correlate positively with these belief items, while importance ratings of social features did not tend to correlate with these belief items. This suggests that beliefs about gender being binary, or biologically essentialised, tend to be related to estimations of how much people use biological features to categorise others by gender, but not how much they use social features.

Table 1

Spearman's Correlations between Importance Ratings of Features and Gendered Belief Items

Belief	Importance rating				
	Biological features of the body	Biological features of the face	Social features on the body	Social features on the face	Social expression
"Someone born male is a man, regardless of whether he is wearing a dress"	rho = .20* *p* = .012	rho = .22* *p* = .005	rho = .05 *p* = .571	rho = .11 *p* = .171	rho = -.11 *p* = .184
"Someone born female is a woman, regardless of whether she is wearing a suit"	rho = .15 *p* = .058	rho = .22* *p* = .006	rho = .06 *p* = .460	rho = .11 *p* = .157	rho = -.11 *p* = .182
"When I first meet someone, I almost always think of them as men or women, and rarely as non-binary"	rho = .33* *p* < .001	rho = .25* *p* = .001	rho = .17 *p* = .037	rho = .13 *p* = .093	rho = -.19* *p* = .015

Discussion

In Study 1b, we aimed to a) investigate importance ratings of social and biological cues of gender by category, and b) explore how these importance ratings may be related to beliefs about gender.

For our first aim we found, as expected, that participants rated both categories of biological features (facial, bodily) as more important than both corresponding categories of social features (facial, bodily) of gender. However, we also found that social features were rated almost as highly as biological features (effect sizes were small), showing that participants also view social features as relatively important. Our findings are in line with those of Study 1a, and further suggest that participants may view gender as partly biological and partly social.

For our second aim, we largely found that beliefs about gender being binary and biologically essentialized were related to importance ratings of biological features, but mostly not related to importance ratings of social features. Our findings suggest that binary beliefs about gender may be related to overestimations of one's use of biological features to categorise others. This aligns with the binary, biologically essentialized conceptualization of gender (Saguy et al., 2021). Estimations of how much people use social features, on the other hand, are mostly not affected by these beliefs. This indicates that people may find social features important for gender categorisation, regardless of their ideological beliefs about gender.

Across both Studies 1a and 1b, we had investigated people's explicit ideas about using biological and social features during gender categorisation. Next, we wanted to additionally explore their more implicit ideas by looking at their behaviour during a gender categorisation task. As such, in Study 2a, we aimed to explore participants' gender categorisations based on biological and social

cues. For the biological cues, we decided to show participants images of faces, a type of stimulus commonly used in traditional gender categorisation tasks (Domen et al., 2020; Freeman et al., 2010; Kecskés-Kovács et al., 2013; Mouchetant-Rostaing et al., 2000). For the social cues, we chose to show participants images of shoes. Shoes are a form of gender expression that participants had rated as highly important in Study 1a, and which is typically not highly affected by biological body shape (male and female feet typically only differ visibly in size while other differences are not easily seen with the naked eye; Luo et al., 2009). Thus, we wanted to explore what would happen when the information in the social cue differed from the information in the biological cue, for instance, when somebody's biological facial features suggest that they are male, but their clothing is feminine. In line with results from Studies 1a and 1b, we anticipated that participants would base their gender categorisations more on the biological cue than the social cue, overall (especially given neuroscientific evidence that faces are particularly attention-grabbing; Fusar-Poli et al., 2009).

More interestingly, we also wanted to investigate under which conditions the social cue may override the biological cue in people's decision-making process. As such, in Study 2a, we explored the role of social context, whereby participants were placed in one of two conditions, which we based on feasible real-world scenarios. In one condition, participants were asked to categorise the other person by imagining that they were addressing them as "Mr." or "Mrs.", thus their categorisation was part of an interpersonal interaction. In the other condition, participants were asked to imagine addressing the other person by their first name, and it was implied that their gender categorisation was 'private' (thus, the other person would not know what gender the participant thought they were). We expected social cues to override biological cues more often in the interpersonal condition than

the private condition. This is because social cues indicate how people want to present themselves in interactions with others, and they may thus be seen as more relevant by participants in interpersonal situations than in private ones.

Lastly, in order to compare participants' behaviour with their own estimations of their decision-making process, we asked participants to indicate how much they thought they based their decision on the face or the shoe.

Study 2a

Method

Participants

Our final sample size was $N = 85$. Participants were mostly female (85.88%), cisgender (100.00%), heterosexual (88.24%), and White (92.94%). We found no difference in results between people of different demographics. For more information, including power analysis, see Supplementary Information.

Materials

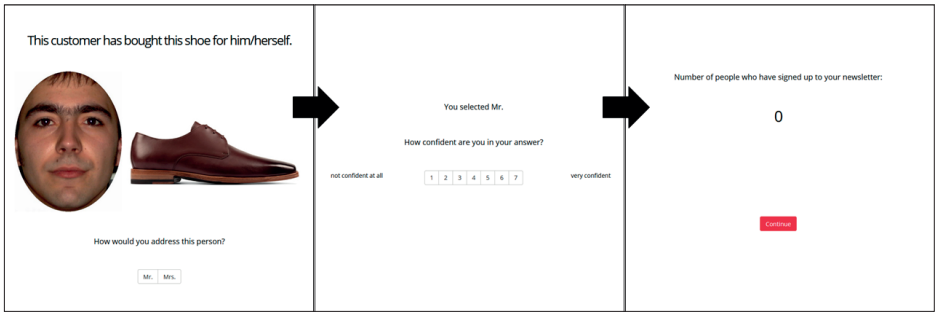
Gender categorisation task. Participants were asked to imagine that they had started working in customer service for a shoe sales company. They imagined that their job was to video call customers and convince them to sign up to a newsletter, and that their performance would be assessed based on how many customers signed up to the newsletter. In each of the 180 trials, participants viewed a unique combination of a morphed face image and a shoe image and told that the customer (depicted in the face stimulus) had indicated buying this shoe for him/herself. Participants were asked to take a break every 45 trials (i.e., there were a total of three breaks), to avoid screen fatigue. There was no time limit for each trial. There were

two conditions, the interpersonal and the private conditions, each with a slightly different sequence of trials (see below).

Interpersonal condition. Below the images of the face and the shoe, participants were asked how they would address each customer: as "Mr." or "Mrs." (and then their last name). This was thus the condition where we implied that participants were categorising the customer as part of an interpersonal interaction. On the next screen, participants indicated how confident they were in their gender choice (7-point Likert scale, 1 = "not confident at all", 7 = "very confident")[4]. At the end of each trial, they were given bogus, pseudo-randomised feedback regarding whether the participant had signed up to the newsletter or not. We thought that positive feedback in some trials might suggest to participants that they had made the 'correct' gender choice (i.e., in line with the customer's own preferred title), thereby strengthening the idea that there was something at stake when they were making their choices. For an example of a trial, please see Figure 3.

Figure 3

Example of a Trial in the Interpersonal Condition

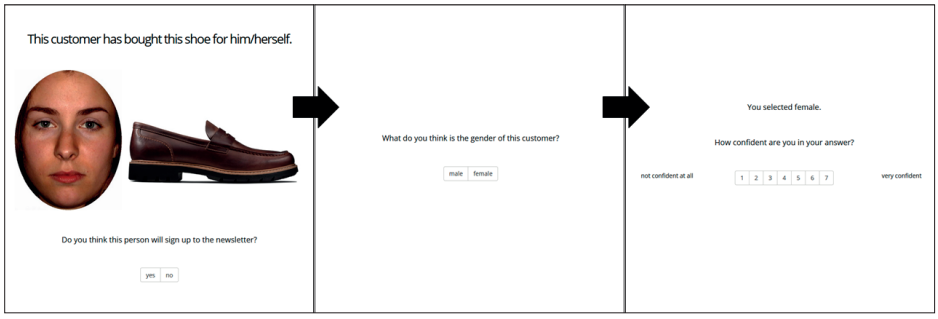


Private condition. Below the images of the face and the shoe, participants were asked whether they thought this participant would sign up to the newsletter. Next, participants were asked what they thought the gender of the customer was ("male" or "female"). In this

condition, participants were told that they would address the customer by first name. We thus implied that the customer would not be aware of the participant's gender categorisation, and thus their categorisations of the customer were private. At the end of each trial, participants indicated how confident they were in their gender choice (7-point Likert scale, 1 = "not confident at all", 7 = "very confident"). For an example of a trial, please see Figure 4.

Figure 4

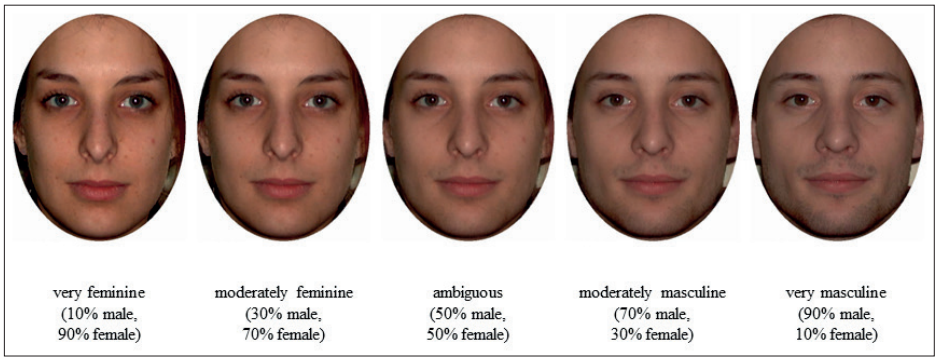
Example of a Trial in the Private Condition



Face stimuli. A total of 20 face stimuli were taken from Huart et al. (2005). There were five faces which were 10% male and 90% female (very feminine), five which were 30% male and 70% female (moderately feminine), five which were 50% male and 50% female (ambiguous), five which were 70% male and 30% female (moderately masculine), and five which were 90% male and 10% female (very masculine). For an example of faces, please see Figure 5.

Figure 5

Example of Face Stimuli



Shoe stimuli. We selected 24 images of shoes, taken from the Clarks digital shoe shop (https://www.clarks.co.uk/), which we had piloted among 10 student participants. In the pilot, we asked participants to what extent they thought each shoe was typical for women or men ("designed for women/men", "bought by women/men", "suitable for women/men", and "feminine/masculine"; on a 7-point Likert scale, 1 = "Not at all", 7 = "Very much"). We selected three typical men's shoes, which pilot participants had rated typical for men ($M \geq 5.63$), but not typical for women ($M \leq 2.48$). We selected three typical women's shoes, which participants had rated typical for women ($M \geq 6.13$), but not typical for men ($M \leq 2.28$). Lastly, we selected three ambiguous shoes, which participants had rated relatively equally typical for men and women ($M^{difference} \leq 1.48$). Our final shoe stimuli thus consisted of three masculine shoes, three feminine shoes, and three ambiguous shoes. For an example of each shoe, please see Figure 6.

Figure 6

Examples of Feminine, Masculine and Ambiguous Shoe Stimuli

Gender categorisation estimates. To measure how much participants estimated having used each stimulus for their decision making during the categorisation task, we showed them the following items: "I used the picture of the face in my gender categorisation" and "I used the picture of the shoe in my gender categorisation" (7-point Likert scale; 1 = "Not at all", 7 = "Very much"). We analysed these items separately from each other.

Procedure

The procedure for ethical approval, participant recruitment and informed consent were the same as in Study 1a. Participants completed this study on Gorilla Experiment Builder (Anwyl-Irvine et al., 2018). Participants were randomly assigned to one of the two conditions (interpersonal or private) and completed the gender categorisation task. Next, participants filled in manipulation check items, items about task engagement, threat of misgendering, and how acceptable they find non-normative clothing.[5] Next, they answered the items about gender categorisation awareness. Following this, they answered demographic questions regarding their gender identity, sexual orientation, and ethnicity. Lastly, they answered a question with which we checked whether they had participated in another of

our studies within the same project.[6] Participants were debriefed and received study credit for their participation.

Results

Context-dependent gender categorisation

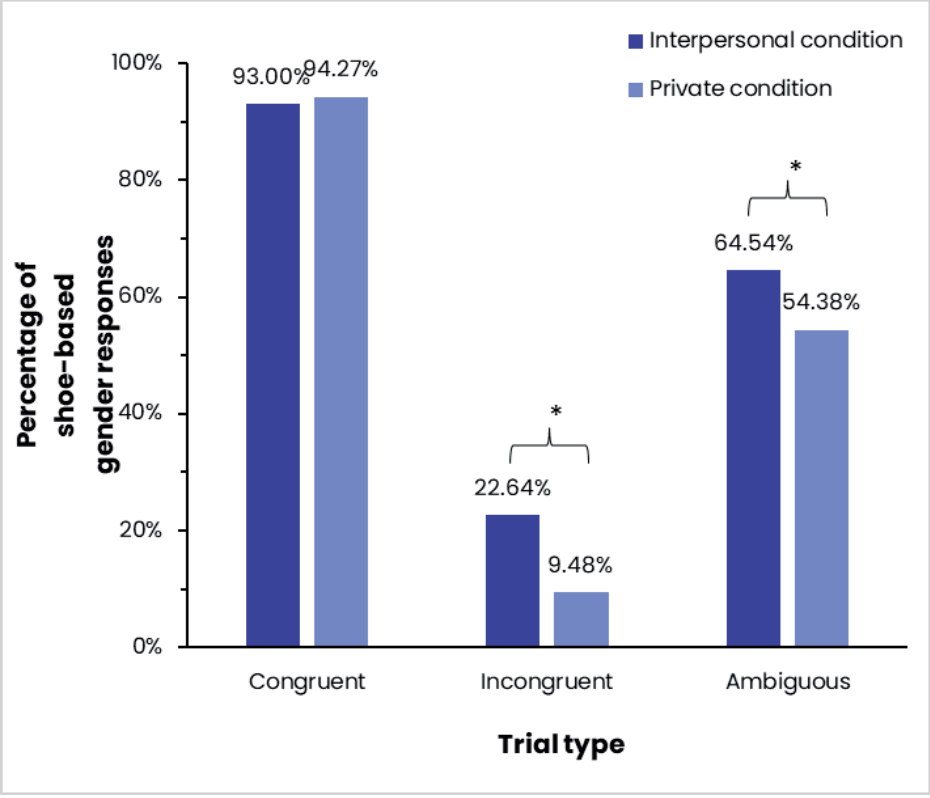
We wanted to investigate differences in gender categorisation between conditions, in order to see whether the *context* in which one is categorising someone's gender affects gender categorisation. For the purpose of analysis, we created a variable, 'shoe-based gender response'[7], and coded all gender responses that were in line with the gender of the shoe (e.g., participants chose "Mrs." when shown a feminine shoe) as 'shoe-based'. Conversely, we coded all trials where gender responses were not in line with the gender of the face (e.g., participants chose "Mr." when shown a feminine shoe, not including ambiguous shoes) as 'not shoe-based'. Next, we created a variable, 'congruence', and categorised trials in which both the face and the shoe were (moderately or very) feminine/masculine as 'congruent', trials in which the face and the shoe had opposing gender information (e.g., feminine shoe with masculine face) as 'incongruent', and all other trials which included a gendered shoe (e.g., ambiguous face with feminine shoe) as 'ambiguous'.

We expected that participants would make more shoe-based responses (i.e., responses in line with the social feature) in the interpersonal condition than in the private condition. Specifically, we expected this difference to be present in incongruent and ambiguous trials, where the face and the shoe held different gender information. In order to test this, we ran a series of three chi-square tests (IV = condition, DV = shoe-based gender response), one per trial type (congruent, incongruent, ambiguous). We found a significant difference between conditions in both the incongruent ($\chi^2(1) = 132.10$, $p < .001$, $\varphi = 0.18$), as well as the ambiguous

trials ($\chi^2(1) = 22.12$, $p < .001$, $\varphi = 0.10$), but not the congruent trials ($p = .075$), such that participants made more shoe-based gender responses in the interpersonal condition than in the private condition, as expected (see Figure 7). This implies that participants in the interpersonal condition gave more weight to the social cue (the shoe) and thus less weight to the biological cue (the face), than participants in the private condition.[8]

Figure 7

Percentages of Shoe-Based Gender Responses per Trial Type and Condition



Estimates of gender categorisation

To test whether participants' estimations of how much they used the face or shoe stimuli for their categorisation differed per

condition, we ran two independent-samples Mann-Whitney U tests, one for each item. We found that, in the interpersonal condition, participants reported using the face less than in the private condition (M^1 = 6.11, SD_1 = 0.91, N_1 = 45, M_2 = 6.63, SD_2 = .67, N_2 = 40, U = 1208.50, p = .003, r = 0.33), in line with their actual behaviour in the task. However, we found no significant difference between conditions in how much participants reported having used the shoe for their categorisation (M = 3.21, SD = 1.58, p = .139), even though their behaviour suggested that they did use the shoe more. This suggests that participants in the interpersonal condition may not have realized that they had given more weight to the shoe stimulus than participants in the private condition. Across both conditions, participants reported using the face (M = 6.35, SD = 0.84) significantly more than the shoe (M = 3.21, SD = 1.58), as revealed by a related-samples Wilcoxon signed rank test (Z = 24.00, p < .001, r = 0.82). This suggests that participants thought the biological cue (the face) was more important than the social cue (the shoe) for gender categorisation, regardless of condition. This is also in line with the low shoe-based gender responses during the task.

Discussion

Our aims for Study 2a were a) to explore participants' gender categorisations based on a biological and a social cue, b) to investigate whether an interpersonal context would affect gender categorisations, as compared to a private context, and c) to compare participants' estimations of their own decision-making with their behaviour during the gender categorisation task. For our first two aims, we found that, while participants tended to base their decisions less on the social cue than the biological cue, there were important differences between social contexts. Specifically, we compared two conditions: 1) a private context, in which participants categorised the other person with seemingly no social consequences, and 2) an interpersonal context, in which participants

categorised the other person for the purposes of an interpersonal interaction. We found that, in the interpersonal context, participants tended to base their decision more on the social cue than in the private context, as expected. We think this is because people use social cues (e.g., clothing) as a form of gender expression, making the social cue more relevant in a context where you have to use information to know how to correctly address someone. These findings suggest that we should understand gender categorisation as a tool for social communication, and not only a processing of information on a cognitive level.

For our final aim, we found that participants' estimations of how much they used the biological cue differed between the private and interpersonal conditions, but their estimations of how much they used the social cue did not. Thus, though participants' behaviour suggested that they used the social cue more in the interpersonal condition, they may not have been aware of this. This was in line with our expectations, since people are often not aware of their own categorisation processes (Rule & Sutherland, 2017) and this may be even more the case for a highly debated topic like gender (Cooper, 2019).

Our findings in Study 2a indicated that a) the social cue can override the biological cue in participants' decision-making during a gender categorisation task, in the right social context, and b) participants may be somewhat unaware of this. While in Study 2a we investigated gender categorisation in a less biologically essentialized, and more nuanced way than many previous studies, the gender choices we provided participants were still binary. In Study 2b, we wanted to additionally include conditions in which participants could choose from more than two gender categories, which is becoming more common in modern society (e.g., Davis, 2023). Given that we had found a difference in gender categorisations between the interpersonal and private contexts in Study 2a, we anticipated that we may also find a difference between the

binary and non-binary contexts in Study 2b.

Furthermore, in Study 2b, we wanted to add another measure to the gender categorisation task, namely attention. Specifically, we wanted to explore how much attention they paid to the face and the shoe while making their decision, to see whether this was in line with their behaviour. Doing so could provide us with further insight into the processes underlying participants' decision-making.

Study 2b

Method

Participants

Our final sample size (N = 167) consisted of mostly students (66.47%), who were female (71.26%), cisgender (97.00%), heterosexual (64.67%), and White (65.87%). We recruited 37.72% of participants via the Utrecht University and KU Leuven student participation systems, while the remaining 62.28% of participants were recruited on Prolific (https://www.prolific.co/). We mostly recruited participants living in the Netherlands (41.32%) and Belgium (46.11%), to replicate our previous study samples. We checked for differences between participants of different demographics, which are beyond the scope of this chapter and can be found in the Supplementary Information.

Materials

Gender categorisation task. We used the same categorisation task as in Study 2a, with four conditions, rather than two.

Interpersonal, binary condition. This condition was the same as the 'interpersonal condition' in Study 2a.[9]

Private, binary condition. This condition was the same as the 'private condition' in Study 2a.

Interpersonal, non-binary condition. This condition was the same as the 'interpersonal, binary condition', except that participants had more than two gender choices: 'Mr.', 'Mrs.', 'Mx.', and 'I don't know'.

Private, non-binary condition. This condition was the same as the 'private, binary condition', except that participants had more than two gender choices: 'male', 'female', 'non-binary', and 'I don't know'.

Eye-tracking. We tracked participants' eye fixations on the screen using a webcam-based eye-tracker, Webgazer.js (Papoutsaki et al., 2016). The face stimulus was located entirely on the left side of the screen, while the shoe stimulus was located entirely on the right side of the screen. Eye-tracking data from Gorilla Experiment Builder (Anwyl-Irvine et al., 2018) records the time (in ms) that a participant fixated on each side of the screen: left and right. From this we extrapolated the percentage of time that participants spent looking at either the face (left) or the shoe (right), relative to each other. This gave us an indication of participants' attention to each stimulus.

Gender categorisation awareness. We included the same self-report items as in Study 2a.

Procedure

The procedure for ethical approval and informed consent were the same as in Study 1a. Participants were unable to participate on their phone. Participants were redirected from either the universities' student participation systems, or from Prolific, to the experiment on Gorilla Experiment Builder (Anwyl-Irvine et al., 2018). Participants were asked to fill in some questions about their set-up, to check for differences in eye-tracking results depending on lighting (natural or artificial, bright or dark), device (laptop or tablet), device placement (on surface, on lap), and whether they wore glasses.[10] Next, participants were randomly assigned to one of the four conditions.

They completed an eye-calibration where they had to fixate on five dots (one in each corner of the screen and one in the middle) twice for 10 seconds, before doing the gender categorisation task. After each break within the task, participants had to recalibrate their eyes to continue. After completion of the task, participants filled in the self-report measures.[11] in the same order as Study 2a. Following this, they answered demographic questions regarding their gender identity, sexual orientation, and ethnicity. There were two attention checks throughout the surveys. Lastly, participants were debriefed and received monetary compensation for their participation.

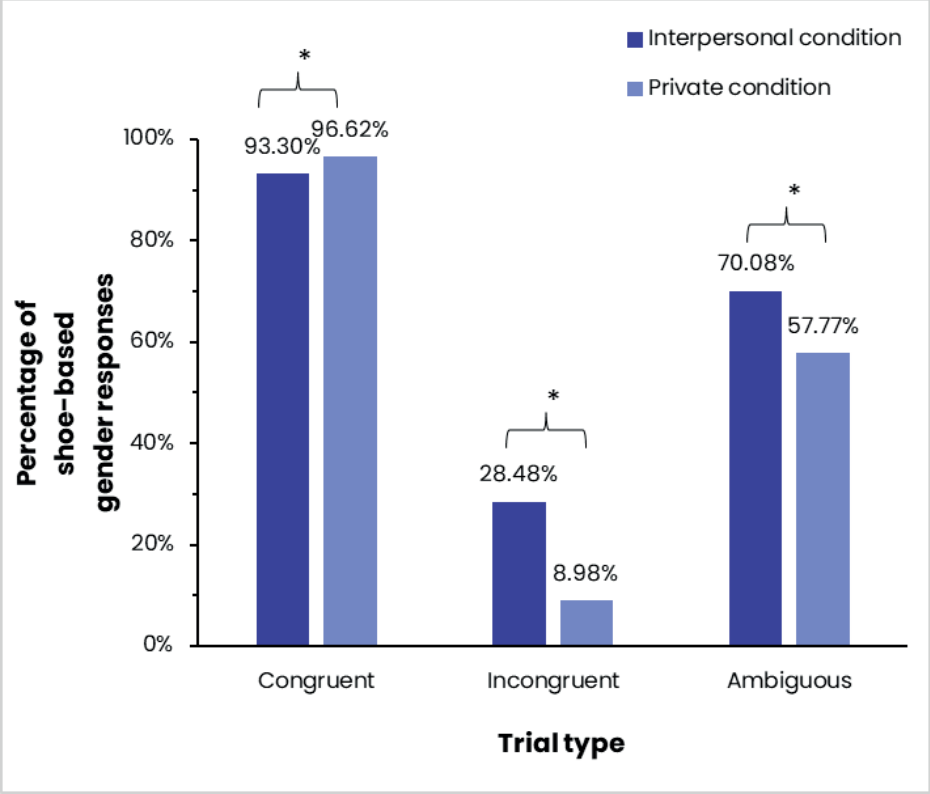
Results

Gender categorisation dependent on private or interpersonal context

Just like in Study 2a, we wanted to test whether the percentage of socially based (i.e., shoe-based) gender responses differed between the interpersonal and private conditions. We thus ran a series of three chi-square tests (IV = condition, 2 levels; DV = shoe-based gender response), one per trial type (congruent, incongruent, ambiguous), the same as in Study 2a (see Figure 8).

Figure 8

Percentages of Shoe-Based Gender Responses per Trial Type and Interpersonal or Private Condition, Collapsed over Binary and Non-binary Conditions



As expected, and replicating our previous results, we found significant effects of condition on percentage of shoe-based gender responses in incongruent ($\chi^2(1) = 417.44$, $p < .001$, $\varphi = 0.25$) and ambiguous trials ($\chi^2(1) = 55.85$, $p < .001$, $\varphi = 0.13$). In both incongruent and ambiguous trials, the percentage of shoe-based gender responses was higher in the interpersonal condition, replicating findings from Study 2a.[12] Additionally, we found a significant effect of condition on percentage of shoe-based gender responses in congruent trials ($\chi^2(1) = 44.19$, $p < .001$, $\varphi = 0.08$). Specifically, we found that participants made

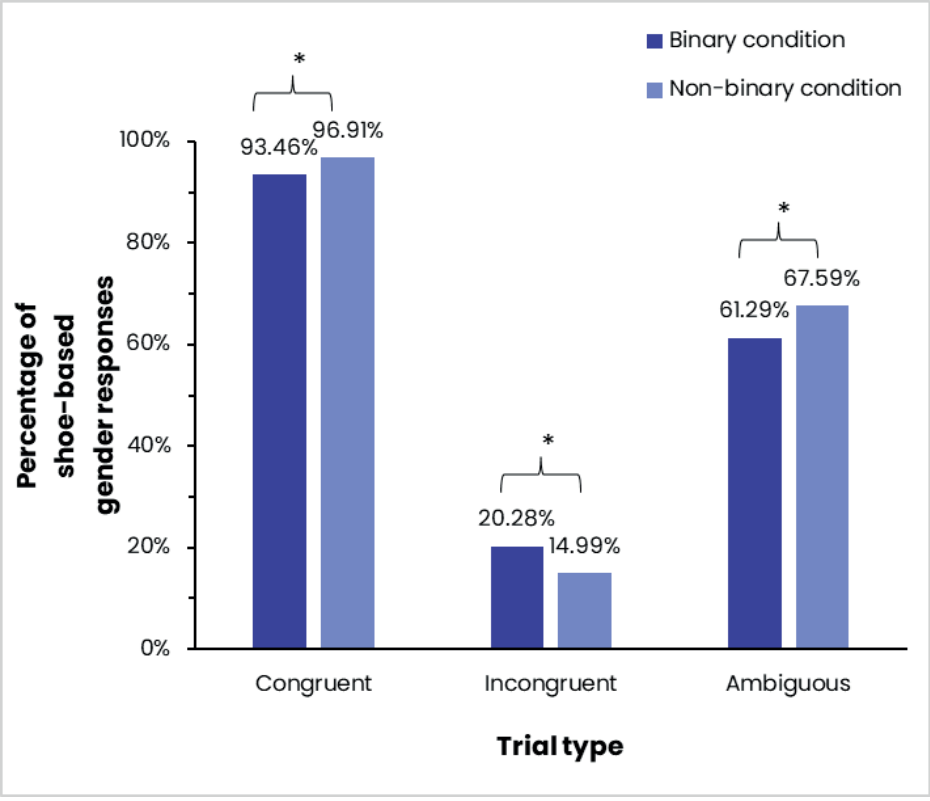
higher numbers of shoe-based (and thereby also face-based, since trials were congruent) responses in the private condition, suggesting a spillover effect from incongruent and ambiguous trials.

Gender categorisation dependent on binary or non-binary context

We wanted to explore whether giving participants more than two gender options to choose from (non-binary condition) would result in different amounts of shoe-based gender categorisations as compared to giving participants only two options (binary condition). We thus ran a series of three chi-square tests (IV = condition, 2 levels; DV = shoe-based gender response), one per trial type (congruent, incongruent, ambiguous; see Figure 9).

Figure 9

Percentages of Shoe-Based Gender Responses per Trial Type and Binary or Non-Binary Condition, Collapsed over Interpersonal and Private Conditions



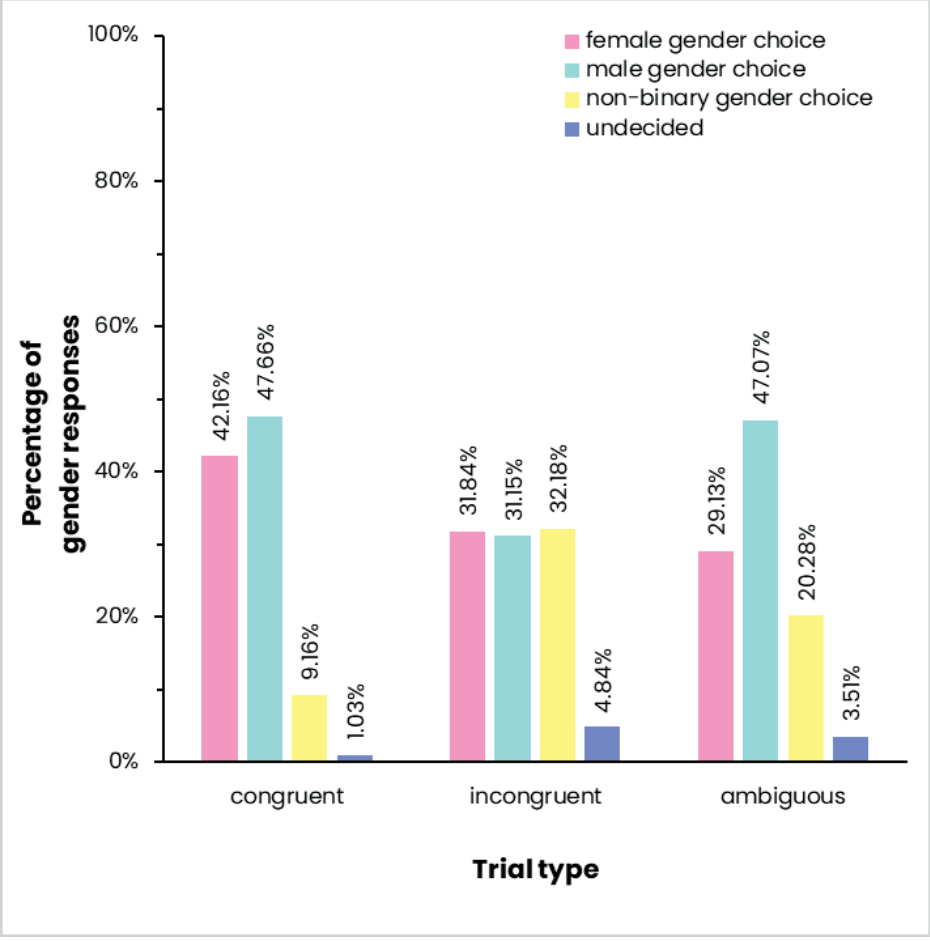
We found a small significant effect of non-binary condition on the percentage of shoe-based responses (and thereby also face-based responses, given that the two stimuli were congruent) in congruent trials ($\chi^2(1) = 47.70$, $p < .001$, $\varphi = 0.08$), suggesting another spillover effect. More importantly, we found a small significant effect of (non-)binary condition on the percentage of shoe-based responses in incongruent trials ($\chi^2(1) = 28.77$, $p < .001$, $\varphi = 0.07$). Specifically, we

found that participants made more shoe-based responses in the non-binary condition than the binary condition. We also found a small significant effect of (non-)binary condition on the percentage of shoe-based responses in ambiguous trials ($\chi^2(1) = 13.95$, $p < .001$, $\varphi = 0.06$). Specifically, we found that participants in the non-binary condition made a higher number of shoe-based responses than participants in the binary condition, opposing the pattern of results for incongruent trials. This suggests that participants in the non-binary condition relied more on the gendered information in the shoe (when coupled with an ambiguous face) than participants in the binary condition.

We wanted to explore why the pattern of results was different in incongruent than in ambiguous trials for the non-binary condition. We suspected that these differences may have been to do with the number of times participants chose the non-binary gender option in each trial type.[13] As such, we ran a chi-square test (IV = trial type, 3 levels; DV = gender response, 4 levels) including only participants in the non-binary conditions, and found differences in type of gender response given per trial type (see Figure 10). The chi-square was significant ($\chi^2(6) = 897.31$, $p < .001$, $\varphi = 0.25$) and post-hoc z scores ($p \leq .050$) revealed significant differences between all conditions. Interestingly, in incongruent trials, participants more often chose the non-binary gender option than in the ambiguous trials. This indicates that participants were more likely to choose a non-binary gender option when shown conflicting gender information (e.g., a masculine face and a feminine shoe) than when shown an ambiguous face and a gendered shoe. This is likely part of why we found a different pattern of results for trials in which there was conflicting information (incongruent trials) versus trials in which there was ambiguous information.

Figure 10

Percentages of Gender Choices Selected per Trial Type



Eye-tracking: Relationship between condition and attention to each stimulus

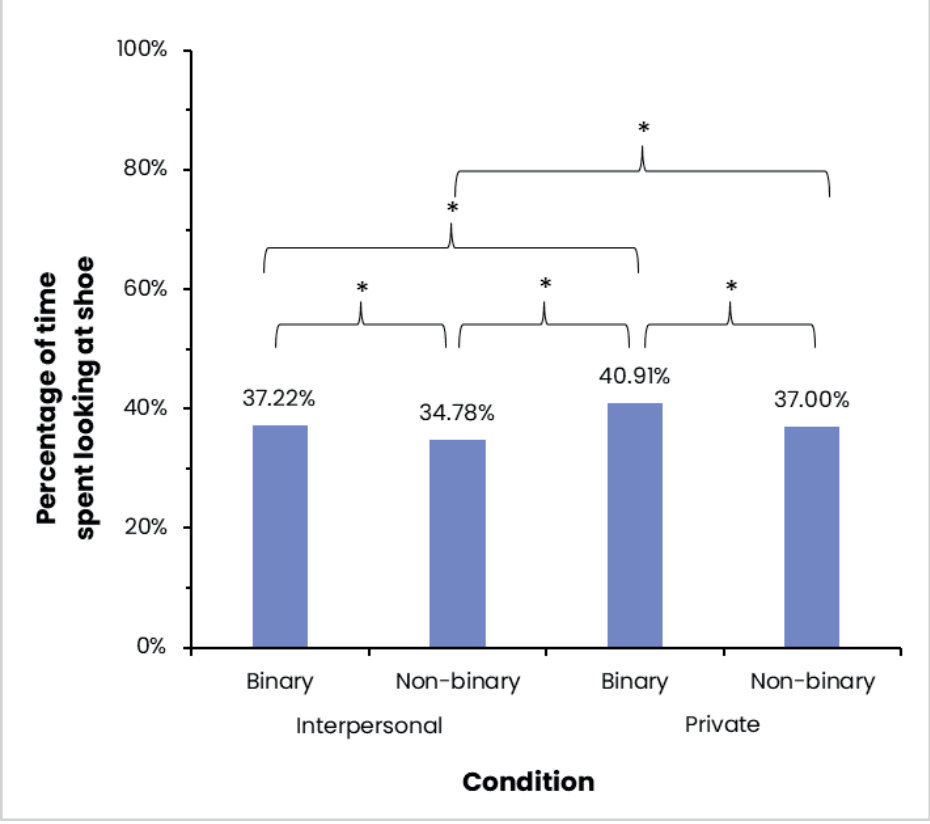
We explored whether participants' attention to the social and biological cues of gender matched their behaviour, in that they paid more attention to the social cue (the shoe) than the biological cue (the face) in conditions where they made more socially based gender responses. We thus tested whether the percentage of time looking at

the shoe relative to the face differed between conditions. We chose this, rather than testing for differences in time spent looking at the shoe (in ms) and the face (in ms), to be able to make direct comparisons between behavioural data (which is in percentages) and attention data. Furthermore, we wanted to make sure our results were not confounded by total time looking at both stimuli (in ms), which we had found to differ between conditions.[14]

We thus tested, in a one-way ANOVA, whether the percentage of time spent looking at the shoe (relative to the face) differed between the four conditions (interpersonal binary, private binary, interpersonal non-binary, and private non-binary). We found a significant main effect of condition ($F(3,30051) = 69.55$, $p < .001$, $\eta_p^2 = .01$). A Tukey post-hoc test showed that participants spent more time looking at the shoe in conditions where they tended to make less shoe-based gender responses (see Figure 11). Specifically, we found that participants looked at the shoe the least in the interpersonal, non-binary condition (significantly different from all other conditions, $ps < .001$). Participants looked at the shoe the most in the private, binary condition (significantly different from all other conditions, $ps < .001$). Participants in the interpersonal, binary and the private, non-binary conditions looked at the shoe to an equal extent ($p = .945$). This indicates that, even when participants do not base their decision on the social cue, they spend considerable time looking at it, showing that they do pay attention to it (see Figure 11). We think this further highlights the importance of social cues in gender categorisation.

Figure 11

Percentage of Time Spent Looking at Shoe per Condition



Note. Stars (*) denote significant differences between conditions.

In a factorial ANOVA we checked whether the effect of condition interacted with either the congruence of the trial, the type of face, or the type of shoe, but all interactions were non-significant (*ps* ≥ .221). This suggests that the effect of condition on time spent looking at the shoe (relative to the face) was not affected by the type of trial.[15]

Gender categorisation estimates

To test whether participants reported using the face or the shoe more, and whether this differed between conditions, we ran a related-

samples Wilcoxon signed rank test. This (Z = 587.50, p < .001, r = 0.64) revealed that, across all conditions, participants reported using the face (M = 6.24, SD = 1.03) significantly more than the shoe (M = 4.43, SD = 1.74). This indicates that participants thought the biological cue (the face) was more important than the social cue (the shoe) for gender categorisation, regardless of condition. This is also in line with their gender responses in the task, as well as their attention to each cue as revealed by the eye-tracking.

More interestingly, to test whether participants' estimations of how much they used the face or shoe stimuli for their categorisation differed per condition, we ran two independent-samples Kruskall-Wallis tests, one for each item. These revealed no significant differences between conditions in how much participants reported using the face or the shoe (ps ≥ .157). This suggests that participants may not have been aware that they used the face less or the shoe more in the interpersonal and non-binary conditions.

Discussion

Our aims for Study 2b were a) to replicate findings from Study 2a regarding the effect of interpersonal versus private context on gender categorisation, b) to additionally explore the effect of binary or non-binary context on gender categorisation, c) to explore participants' attention to social and biological cues using eye-tracking, and d) to replicate findings from Study 2a regarding participants' own estimates of their decision-making processes during gender categorisation.

For our first aim, we found, just as in Study 2a, that participants based their gender categorisations more often on the social cue in the interpersonal context than the private context. This was the case regardless of whether the participant was shown binary or non-binary

gender options to choose from. This highlights the robustness of the effect of interpersonal context on gender categorisation based on both social and biological cues.

For our second aim, we found a small and less robust effect of non-binary context, as compared to binary, on gender categorisation. Specifically, providing participants with non-binary gender options seemed to be related to a higher number of shoe-based gender responses in ambiguous trials, and lower numbers of shoe-based gender responses in incongruent trials. When exploring this further, we found that participants tended to select the non-binary gender choice more often in incongruent trials than ambiguous trials, partly explaining these results. It thus seems that participants tend to assume that someone is non-binary more often when provided with conflicting gender information (e.g., masculine face and feminine shoe). As such, the "non-binary" gender option may represent "something in between male and female" in people's minds. On the other hand, when provided with an ambiguous face and a gendered shoe (e.g., feminine shoe), participants tended to rely more on the gender information in the shoe and select a binary gender choice. While these effects were somewhat smaller than the effects of interpersonal context (as compared to private context) on gender categorisation, they do suggest that a non-binary context (i.e., where a non-binary gender option is provided) affects how people categorise by gender. This further highlights how gender categorisation can be based on context and hence, is a social performance besides a cognitive process.

For our third aim, our eye-tracking results revealed that participants looked at the social cue the longest in conditions where they tended to base their decision less on the social cue. This suggests that, even when participants ultimately base their gender categorisation on the biological cue, they are paying attention to the social cue,

perhaps more so than when they base their decision on the social cue. This highlights that people take the social cue into substantial consideration, even though they base their final decision on the biological cue.

Furthermore, just as in Study 2a, we found that participants did not report using the social cue (the shoe) more for their decision-making process in the conditions where they actually did use it more. Additionally, unlike in Study 2a, participants also did not report using the biological cue (the face) less in conditions where they did, in fact, use it less. However, our eye-tracking results also provide us with more insight as to why this might be, given that we found that participants tended to look at the social cue more (and thus the biological cue less) in conditions where they tended to base their decision on the social cue less (and thus more on the biological cue). As such, participants who based their decision less on the social cue may judge that they used the social cue to come to their decision, since they paid attention to it (for an overview of results, see also Figure 12).

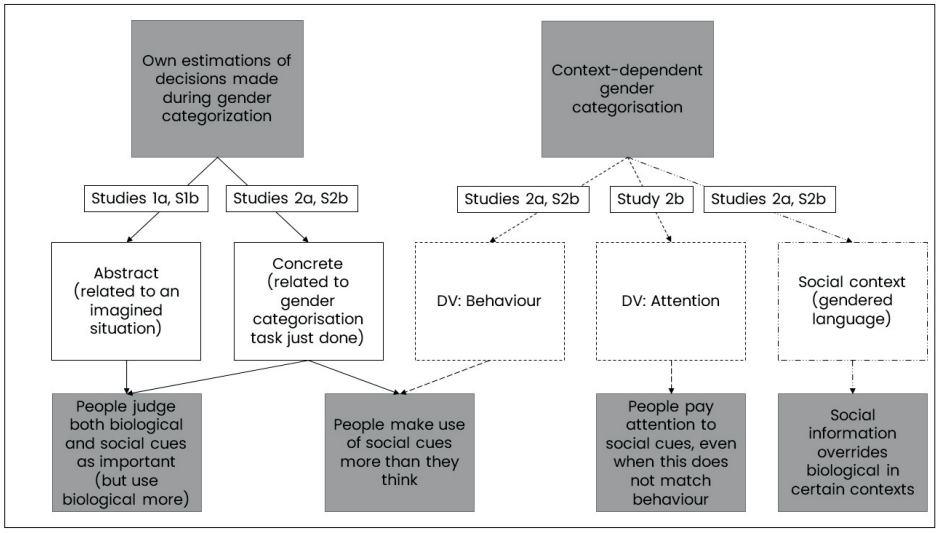
General Discussion

Across four studies, we investigated to what extent people use biological (e.g., facial structure) and social (e.g., clothing) information to decide the gender of people they encounter. Doing so provided us insight into what participants consider to be defining characteristics of people's genders and thus engage with the current public debate around gender (Cooper, 2019). In the first two studies, we investigated gender categorisation in a more abstract way, by asking participants to imagine meeting a new person and to reflect on how they decide the gender of that person. In the latter two studies, we did so in a more concrete way, by asking participants to categorise people by gender, based on a picture of a face and a shoe. Throughout four studies, our central aim was to investigate how people use biological and social

information to categorise by gender, both explicitly (self-report, all studies) and implicitly (behaviour in a gender categorisation task, Studies 2a and 2b; attention to stimuli, Study 2b). Furthermore, to better understand under which conditions social information may override biological information during gender categorisation, we also manipulated social context based on real-world gender-inclusive language initiatives (Studies 2a and 2b). For a visual representation of the themes of each study, how they relate to one another, and our main findings, please see Figure 12.

Figure 12

Visual Representation of the Research Themes, Dependent Variables and Main Findings in this Chapter



Explicit and implicit measures of gender categorisation

As per our main aim, we were able to gain a better understanding of people's explicit estimations about the type of information (biological and social) they use to understand someone's gender and compare this to their behaviour and attention during a

gender categorisation task, which represents a more implicit measure. Specifically, we found that, overall, participants judged biological information to be more important for determining someone's gender than social information. We found this when we asked participants to rate specific features (e.g., presence or absence of a beard; Study 1a), when we asked participants to rate categories of features (e.g., biological features of the face; Study 1b), as well as when we asked participants to reflect on their own behaviour during a gender categorisation task (Studies 2a and 2b). This was in line with our expectations and the societal narrative, in which gender is still largely portrayed as a biologically essentialized concept (Fine, 2017). However, we also found, across all studies, that participants did not rate social cues of gender as *unimportant*, showing that people to some extent view social expression (e.g., clothing) as another defining characteristic of gender. This highlights that the way people use information to determine someone's gender may be somewhat more nuanced than the polarized societal debate around gender would have us believe (Cooper, 2019).

A strength of the studies in this chapter is that we investigated how people use biological and social information for gender categorisation, both explicitly and more implicitly. Our results also show why it is important to investigate both. Specifically, in Studies 2a and 2b, we found that participants used social information more in certain conditions than others but did not *report* doing so. This highlights that people are not always aware of their own thought processes (Rule & Sutherland, 2017). Furthermore, our eye-tracking findings (Study 2b) also provided further insight into participants' thought processes during gender categorisation. Specifically, we found that in conditions where they ultimately based their decision on the biological information, they looked at the social information for longer.

We thus show that investigating gender categorisation from different angles, using different measures, can provide a more complete picture of how people process gendered information about a person.

Gender categorisation and social context

To give us further insight into the conditions under which social information may override biological information, we also manipulated social context during the gender categorisation task. Specifically, we were inspired by real-world gender-inclusive language initiatives (Davis, 2023; Merriam-Webster, n.d.) and wanted to know how those may affect gender categorisation. As such, in Studies 2a and 2b, participants either imagined addressing people in a gendered way (using titles: 'Mr.' or 'Mrs.') or addressing them by their first name. These are realistic scenarios, since we often have to make a decision about someone's gender in order to be able to address them formally in daily life. Furthermore, in Study 2b, we added two further conditions, in which participants could additionally choose from non-binary gender options (e.g., the gender-neutral title 'Mx.'). We are currently seeing an increase in gender-inclusive language initiatives, such as gender-neutral pronouns (Sendén et al., 2021), highlighting a shift in the social narrative surrounding gender and gendered language. While it is not yet common to use non-binary language everywhere, these kinds of initiatives have been gaining considerable media attention (e.g., Hawkins, 2022; McGann, 2017; Swerling, 2023), making this a timely topic that participants have also likely heard about. We found differences in gender categorisations between the different social contexts, whereby social information more often overrode biological information when participants had to address the other person in a gendered way, or when we provided them with non-binary gender options. This highlights that people may process social and biological information differently depending on the situation they are in.

Limitations

While the studies in this chapter explored gender categorisation in a novel way, beyond the binary (and thus biologically essentialized) narrative, the research presented here was only a start. There were thus a number of limitations that should be addressed in future research.

First, throughout our studies, we measured participants' own estimations of how they categorise others by gender and their actual behaviour during a gender categorisation task but did not include all conditions in all our studies. For instance, while in our latter two studies we manipulated social context and investigated how this affected participants' behaviour, we did not manipulate social context in our first two studies (which were self-report). Thus, we were missing some comparisons between explicit and implicit measures depending on different social contexts.

Second, and relatedly, in our latter two studies we compared participants' behaviour in a gender categorisation task with their own retrospective reflection of the type of information they used during the task. We found that participants' reflections did not match their behaviour. However, we investigated this in a between-subjects design. If we had investigated this in a within-subjects design (where participants would have been in both/all conditions), participants may have been better able to judge how they make decisions in different contexts (because they would have been able to compare their choices across conditions).

Third, while we asked participants about many different biological and social features in Studies 1a and 1b, in our latter two studies (2a and 2b) we only used one type of biological cue (faces) and social cue (shoes). Comparing many different types of cues was beyond the scope of these studies but would be useful for testing the robustness

of our effects. Specifically, it is well-documented that faces are treated preferentially in the brain (Fusar-Poli et al., 2009), meaning other types of biological cue (for instance, bodies or body parts) may be more easily overridden by a social cue. Similarly, some social cues may be more or less salient than shoes. We selected shoes as our social cue because participants in Study 1a had indicated finding this the most important feature, and because feet shapes vary less between biological sexes (Luo et al., 2009) than other body parts.

Last, while the use of eye-tracking was a strength of this study, as it allowed us to explore participants' attention to different stimuli in different conditions, the type of eye-tracking we used may have been a limitation. Due to the covid pandemic and having to conduct our experiments online, we used an online eye-tracking measure (Anwyl-Irvine et al., 2018), which was noisier and somewhat more limited than an eye-tracker in a more controlled environment, such as a laboratory. This restricted what we could do with the data, such that we only explored which stimulus participants looked at for longer, overall. Eye-tracking in a more controlled environment may help researchers investigate eye-movements between stimuli, which could give an indication of how they are processing the gendered information. For instance, many eye movements between stimuli could indicate that participants are relying less on their visual working memory and are thus perhaps more unsure of how to use the visual information (Somai et al., 2020).

Implications

There are both theoretical, as well as practical, implications of the research presented in this chapter. Our findings reflect a change in the gender narrative, which we currently see happening around us (Pun et al., 2017), whereby the binary conceptualization of gender is being brought to question. We show that gender categorisation is adaptive depending on social context, and that both biological and

social information are used when categorising others by gender. This has theoretical implications for how we understand gender as a social category, and how we should study gender in social psychology, that is, beyond the binary conceptualization that is still most common in research as of date.

Furthermore, we have shown that participants use social and biological information differently to determine people's genders depending on social context, whereby our contexts were inspired by real-world gender-inclusive language initiatives. We have seen an increase in gender-inclusive initiatives in society (Pun et al., 2017), fueled by the societal debate around gender conceptualization (Cooper, 2019). Our results suggest that real-world gender-inclusive policies can affect how we perceive and interact with others in daily life.

Conclusion

The *gender binary* is the view that humans all fit into one of two categories, men and women, that these categories are biologically meaningful, and that this view is natural and unmalleable (Hyde et al., 2019). Across four studies, we showed that participants do not view biological information as the only important factor in determining someone's gender; rather, they also view social information (e.g., clothing) as important. Furthermore, we show that their gender categorisations may be adaptive to the social context. We also show how explicit and implicit measures can be used together to better understand the underlying processes of gender categorisation. We conclude that it is important for future social psychological research to study gender beyond the binary view, and for policymakers to consider that gender may not be either biological or social, but both.

Footnotes

[1] We also performed all tests excluding participants who had participated in a previous study of the same project but found no different results. We thus decided to include them.

[2] We also included an item ("When I meet someone new, I almost always know their gender just from looking at them") as an assumption check, which was passed; the results of this can be found in the Supplementary Information.

[3] We also performed all tests excluding participants who had participated in a previous study of the same project but found no different results. We thus decided to include them.

[4] Confidence data were used to see how engaged participants were in different conditions. For brevity, we have included the results of this data in the Supplementary Information.

[5] For brevity, the results of these can be found in the Supplementary Information. Importantly, we found our manipulation checks to be successful, and that participants were engaged during the task.

[6] We also performed all tests excluding participants who had participated in a previous study of the same project (N = 9) but found no different results. We thus decided to include them.

[7] We also created a variable, 'face-based gender response', and coded all gender responses that were in line with the gender of the face (e.g., participant chose "Mrs." when shown a feminine face) as 'face-based', similarly to the 'shoe-based gender response' variable. We conducted all of the same analyses with the 'face-based' variable and found the same results as for the 'shoe-based' variable. For the purposes of conciseness of the chapter, we have included those analyses in the Supplementary Information.

[8] We also checked for differences in gender categorisation between trials with different faces, to see how the type of face we showed affected participants' gender choices. We found that participants tended to default to the male gender choice more often than the female gender choice, particularly for ambiguous faces. These results are beyond the scope of this chapter but can be found in the Supplementary Information.

[9] Data regarding how confident participants were in their gender choice can be found in the Supplementary Information, just as for Study 2a. As an additional measure of confidence, we also explored how long participants looked at the stimuli before making their decision, since longer duration would suggest more hesitance. For brevity, these results can also be found in the Supplementary Information.

[10] Descriptive statistics of these can be found in the Supplementary Information. We only included participants with an acceptable set-up for the eye-tracking. All other participants were prevented from taking part.

[11] Data about manipulation checks, task engagement, threat of misgendering and acceptance of non-normative clothing can be found in the Supplementary Information. The manipulation checks were passed, participants were found to be engaged, and we found no interesting relationships between task responses and gender beliefs.

[12] As in Study 2a, we also explored how different types of faces affected participants' gender categorisations and replicated the finding that participants tended to default to the male gender choice more than other gender choices (particularly for ambiguous faces). These results are beyond the scope of this chapter but can be found in the Supplementary Information.

[13] For more results regarding number of times participants chose non-binary options, see Supplementary Information. For brevity, and because

of the scope of this chapter, we do not include these in the main text, but we found that participants chose the non-binary gender a substantial amount of times.

[14] These results, regarding total fixation duration, further highlight how confident participants felt in different conditions and how engaged they were, and shows that the manipulation worked as intended. These results are beyond the scope of the chapter but can be found in the Supplementary Information.

[15] Additionally, we also explored participants' attention to each type of face or shoe stimulus, and found, as one would expect based on the literature, that participants looked the longest at stimuli with carried ambiguous gender information. For brevity, these results can be found in the Supplementary Information.

Chapter 7: General Discussion Attempting to Square the Circle?

"Same hair revolution, unisex evolution. Tomorrow, who's gonna fuss?"

Androgynous – The Replacements

Author contributions: Miriam I. Wickham (Writing – original draft; Writing – review and editing); Félice van Nunspeet (Writing – review and editing); Naomi Ellemers (Writing – review and editing)

The central goal of this dissertation was to investigate commonly held assumptions about the gender binary and the societal shift away from a strictly binary system. The *gender binary* is the view that humans all fit into one of two categories, men and women, that these categories are biologically meaningful, and that this view is natural and unmalleable (Hyde et al., 2019). Our society (in the Netherlands and generally in the Westernised world) is set up in a gender binary way. For instance, a binary gender is shown on our official identification documents with few exceptions. However, in recent years, a shift has been occurring whereby more genders are now being recognised legally (Ryan, 2018), in everyday language (Bradley et al., 2019; Gustafsson Sendén et al., 2015), physical spaces that we frequent (Bovens & Marcoci, 2023) and fashion (Kim et al., 2022), to name a few.

There are a number of assumptions that are commonly held about how or why the gender binary is being maintained. Specifically, that the gender binary is natural to humans and thus hard to move away from, or that the gender binary is helpful to us for social categorisation purposes. I have discussed these in the General Introduction to this thesis, including how this dissertation collects evidence to examine the validity of these assumptions. In this chapter, I focus on the generally held assumptions regarding the shift away from the gender binary. In particular, I focus on assumptions that practitioners may hold. I challenge, based on research evidence provided in my dissertation, the notion that implementing gender-inclusive initiatives will have few positive and many negative effects. For more on these assumptions, and particularly for readers who are themselves practitioners, see the Practical Implications section within this chapter.1 There is also a Dutch version of these practical implications in the 'Nederlandse Samenvatting'.

Below, I summarise the most important finding from each chapter, focussing also on what this finding means, practically speaking, for practitioners looking to take steps away from a strictly binary gender system. Next, I dive into the practical and theoretical implications of the findings in this thesis. Last, I discuss the limitations of the studies reported in this thesis and consider directions for future research.

Part 1: Gender-Inclusive Initiatives

In Part 1 of this dissertation, which includes two chapters, I focused on gender-inclusive initiatives, the discussions about them that I was reading in the media and online, and what positive impacts gender-inclusive initiatives might make to society as a whole.

In Chapter 2, I investigated communication about gender-inclusive initiatives from both the perspective of those who engage in online discourse about this topic (the 'writers'), as well as those who may read the online discourse (the 'readers'). One important finding was that beliefs about gender being binary and negative opinions about gender-inclusive initiatives may be perpetuated online by a small, threatened minority (the 'writers'). On the other hand, the people who read these online opinions (the 'readers') may often not agree with them, but also do not feel strongly enough to engage in the discussion themselves. Thus, the online discourse about gender-inclusive initiatives may be skewed to reflect the opinions of a minority who are against such initiatives. In this chapter, I concluded that we may thus not want to give too much weight to the views expressed in online discourse when considering whether or not to implement a gender-inclusive initiative.

In Chapter 3, I went beyond the current focus in the literature on benefits of gender-inclusive initiatives on non-binary and transgender people. Specifically, I examined whether and how such initiatives may also have hidden benefits for cisgender men and women, by reducing

gender stereotypes in the labour market. One of the current obstacles to men entering women-dominated jobs and women entering men-dominated jobs are gender stereotypes about job *suitability* for each gender. In this chapter, I found that gender-inclusive initiatives may impact beliefs about these stereotypes, which is beneficial for all of society because it may lead to decreased gender discrimination in the labour market and beyond.

Part 2: Gender (Self-)Categorisation

In Part 2 of this dissertation, which includes three chapters, I focused on how we currently understand gender and how this may change if the societal narrative around gender shifts away from a binary view. To do so, I investigated both how we see and categorise our own genders as well as those of other people.

In Chapter 4, I investigated how people self-categorise in terms of gender, and whether those self-categorisations better fit a binary or a non-binary model of gender. I also investigated the consequences of identifying in a non-binary way for various social well-being measures, including feelings of societal inclusion, self-esteem, life satisfaction and general affect. Importantly, I found that a large group of participants self-categorised in a non-binary way, meaning they indicated identifying with both men and women, with neither, or identifying with a third gender. All these people also indicated having lower social well-being than the group who self-categorised in a 'binary' way. In this chapter, I concluded that the move away from a strictly gender binary system could benefit a large number of people by aligning better with their self-views, thus increasing their social well-being, particularly their feelings of societal inclusion. Thus, the shift in the gender narrative may be justified in that it represents how many people feel about their own gender and avoids adverse effects on their well-being due to the adherence to a strictly binary gender system.

In Chapter 5, I extended the findings of Chapter 4, regarding gender self-categorisation, by measuring this implicitly using a reaction-time task. Additionally, I explored implicit categorisations of others and related these to people's self-categorisations. Importantly, this data corroborated findings from Chapter 4, such that I found that many people also self-categorise in gender non-binary ways when we measure this using an implicit measure, which is less prone to social desirability. Additionally, I found that the way people categorise others is also not strictly binary, which provided first evidence that we may not struggle to understand others' genders if gender boundaries become more blurred (i.e., less strictly binary). This provided a foundation for the research in Chapter 6.

In Chapter 6, I further delved into how people categorise others by gender, in terms of what biological and social information they use to do so, and how this may be affected by changes in the gender narrative. Importantly, I found that participants very often used social markers of gender, such as clothing, to understand other people's genders. This fits a non-binary idea of gender (whereby gender is socially constructed rather than entirely biological). Furthermore, in this Chapter I asked participants to imagine working for a company which either had a gender binary policy, whereby "Mr." and "Mrs." were used in communication, or a less binary policy, whereby more gender-neutral language was used in communication. I found that the less gender binary policy affected participants' gender categorisations, such that they used more social information for their categorisation. This highlights how a move away from a strictly gender binary system may affect how people perceive and interact with others and gives us insight into what is ahead as we further shift away from a gender binary system.

Practical Implications: Breaking Down Common Assumptions

Since the end of my contract as a PhD student, I have taken a step away from academia and now work as a researcher in the field. In

my work, I speak to practitioners on a regular basis, for instance human resources managers, diversity and inclusion officers, and government officials. When I tell them about the topic of my PhD, there are a number of assumptions that some practitioners tell me about, which I believe my dissertation can provide some evidence against.

First, practitioners are sometimes afraid to implement gender-inclusive initiatives because they fear that their organisation may face backlash. Certainly, the narrative around these initiatives can often be negative, but, as I also show in Chapter 2 of this thesis, the narrative may be skewed because the discourse is dominated by people who feel very negative about gender-inclusive initiatives. There are also many people who do not express their views very forcefully but may feel positive about changes to the gender narrative, in particular people who feel they do not fit into strict binary categories themselves (which constitute a large percentage of people; Chapter 4).

Relatedly, some practitioners express that nobody at their organisation is non-binary, or in need of gender-inclusive policies or initiatives. Alternatively, they may assume that only a tiny minority of people would benefit from such initiatives, and thus their money is better spent in other domains. As I show in Chapter 4 of this dissertation, one may be surprised how many people do not see themselves as strictly binary: estimates vary between a third and a half of people, in other words, this is not a small minority. Furthermore, gender-inclusive initiatives may be beneficial for the well-being of an organisation's employees who don't entirely 'fit' the binary. Moreover, there are hidden benefits of gender-inclusive initiatives for those who do 'fit' the binary (i.e., cisgender men and women), such that they may decrease gender discrimination in organisations by reducing stereotypes about job suitability (Chapter 3). This may help with attracting and retaining more talent.

Third, practitioners sometimes express hesitation to implement gender-inclusive policies or initiatives because they fear that people will find this change difficult, thus costing a lot of time and effort which could be better spent elsewhere. However, as shown in Chapters 4-6, people already do not see gender as entirely binary, even in the current, rather strict, binary system. Specifically, in Chapter 6, I showed that a gender-inclusive organisational policy was not too difficult for participants to adapt to (as shown by their behavioural responses in the task). Furthermore, in Chapter 3, I found evidence of gender-inclusive initiatives affecting people's beliefs almost immediately after the manipulation. This adaptability suggests that people may not have that much trouble with changes to the gender narrative.

In conclusion, moving away from a gender binary system may feel like attempting to square the circle: an impossibly difficult task. However, many of the findings in this thesis challenge this notion. The take-away that I would like practitioners to remember from this thesis is that gender-inclusive initiatives may be beneficial for their organisations, and are likely to have fewer and less severe negative effects than they might fear.

Theoretical Implications

In Part 1, the main theoretical contribution is highlighting how social psychological literature and methods can add to an existing field of study, in this case, gender. There are two unique contributions that are made possible by seeing gender through a social psychological lens, which I describe below.

Specifically, in Chapter 2, I used an understanding of social psychological threat to uncover what may be driving negative online discourse about gender-inclusive initiatives. In social psychology, threat is often seen as the underlying source of resistance to change,

because people tend to find comfort in the current system, even when this system is not beneficial to them (Jost & Van der Toorn, 2012). Specifically, I used a social psychological framework of threat while observing behaviour in online comments sections and found that people who leave negative comments may be communicating their feelings of threat. People do not intend to communicate their feelings of threat; in fact, they may not be aware of their own threat (Seery, 2013). Thus, their hidden communications of threat may not be picked up on without an understanding of social psychological theory (Morgenroth & Ryan, 2021). Importantly, picking up on threat that people may be feeling can give us a better understanding of why they show resistance to change in the gender narrative, and insights into how to reduce threat to decrease resistance (Scheepers & Ellemers, 2019).

In Chapter 3, I used a social psychological lens by thinking about how majoritised group members (in this case, cisgender men and women) may also be positively affected by including the minoritised group (non-binary and transgender people). Specifically, I investigated whether a gender-inclusive initiative may reduce sexist beliefs, which is linked to lower discrimination, as compared to a gender-binary initiative. Much of the research conducted in the broader, cross-disciplinary field of 'gender beyond the binary' has a strong focus on the minoritised group (Gibson & Fernandez, 2018; Richards et al., 2017). This is invaluable, but in order for initiatives to be successful it is also essential to study the majoritised group and how to increase support from them (e.g., Mallett et al., 2008). This focus on the majoritised group can thus form an important social psychological contribution to the field, by strengthening our understanding of how to increase support for gender inclusion.

In Part 2, on the other hand, rather than using a social psychological lens to add to the broader field of gender studies, I added

to social psychological theory by questioning common assumptions about gender. I believe that, as academics, we must be aware of our own assumptions when conducting research and (at least occasionally) challenge them. I describe examples of how I did so below.

One such assumption is that people only identify with their own explicit gender identity, and not others. For instance, if we want to measure the gender identification of someone who has indicated being a woman, we should only ask her how much she identifies with being a woman (e.g., van Breen et al., 2017). In Chapters 4 and 5, I showed that adding extra gender options to questions about self-identification (or trials in a reaction time task), affected our data by showing a much broader diversity in how people self-identify in terms of gender, as compared to asking about fewer (or one) gender categories. This is a contribution to social identity theory (Turner et al., 1987), which is sometimes overly simplified in terms of the group identities studied. Thinking about people, even those who are seemingly "just" men and women, as diverse in terms of gender group identities (or as having multiple group identities, Wong et al., 2022), affects outcome variables such as well-being (Chapter 4) and categorisation of others (Chapter 5). Thus, we need to consider the complexity of identity more in social psychological research.

A second assumption is about how we categorise others by gender, namely that we do so rapidly, automatically, and based primarily on biological information (e.g., facial structure and body shape). We see this assumption in psychological literature in the way that gender categorisation is studied: in a lab, devoid of human interaction, with participants looking at pictures of faces that have been stripped of social markers of gender (e.g., make-up, hairstyle, clothing; Huart et al., 2005). These kinds of experiments are very clean, in the sense that they do not include many distracting features, such

as social markers of gender, which we consider to be confounding. However, they are also far from realistic, since we typically categorise others during interpersonal interaction: in a social context, wearing make-up, clothes and communicating with our body language (i.e., using social gender markers). In daily situations, we may even update our categorisation of someone based on new information. In Chapter 6, I showed that, if you add additional gender markers as well as manipulate the social context in which you measure gender categorisation, this affects the choices participants make when deciding someone's gender. This adds to theory about social cognition (Fiske & Macrae, 2012), meaning how we understand the social groups of those around us, by showing how categorisation of others is affected by additional (non-biological) information.

In conclusion, there are several unique contributions of this thesis to theory. First, this thesis contributes to gender studies theory by using a social psychological lens to understand the current shift in the gender narrative. Second, it contributes to social psychological theory by using a non-binary gender framework to understand how we categorise the self and others by gender.

Limitations and Directions for Future Research

While this dissertation addresses a highly relevant societal topic, provides many practical implications, and enriches gender studies theory, as well as social psychological theory, some limitations need to be pointed out. These limitations also provide ideas for future research. I describe all of these below.

First, while the research reported in this dissertation has important practical implications, a limitation is that much of the research focused on specific psychological mechanisms and effects. For instance, rather than studying the direct effects of implementing

an intervention in an organisation, I studied how participants reacted to hypothetical interventions that were based on real life, among a sample that was not directly affected by this intervention (Chapters 2, 3 and 6). However, I also supervised various Bachelor (unpublished) and Master (Lemmen, 2021) student theses, where we researched how participants felt about real-life interventions. Although these studies are not reported here, they largely offered evidence confirming results I had found with hypothetical interventions reported in this thesis. A direction for future research emerging from this work is to design, implement and test interventions that limit feelings of threat and maximise support among the majoritised group, while increasing inclusion among the minoritised group.

Relatedly, including research populations with a majority of cisgender men and women was a strength of the research in this thesis, because of the importance of majoritised support for initiatives (Mallett et al., 2008). However, this comes with a limitation with regards to less focus on the needs of the minoritised group in the shift away from a gender binary system. A specific focus on the minoritised group was beyond the scope of this thesis, and also comes with sampling challenges of a group that is difficult to access. However, if future researchers wish to design and test gender-inclusive initiatives, it is essential to put the minoritised group at the forefront: the difficulties that they face and their wishes for initiatives that work to be more inclusive of them.

Similarly, the research was conducted in a Westernised context (most participants were European, some were from the U.S.), with a majority of White participants. Gender norms and gender conceptualisation are culturally and ethnically dependent (Stryker, 2008) and thus we cannot extrapolate all findings in this thesis to different cultural contexts, or people of different cultural backgrounds.

Furthermore, while I measured identities that are known to intersect with gender (sexuality, race/ethnicity, disability and others), in-depth research into the effects of having multiple intersecting identities (Wong et al., 2022) was beyond the scope of this research. It is possible that intersecting identities affect how participants conceptualise gender, their opinions about gender-inclusive initiatives or their needs for inclusion during a shift away from the gender binary (Francis et al., 2022). For instance, while we know from extensive literature that 'colour-blind' interventions often do more harm than good (Jackson et al., 2016), this may not be the case for 'gender-blind' interventions (or 'de-gendering', which is a more inclusive term; Morgenroth et al., 2021). However, individuals of different ethnic groups may feel differently about de-gendering than White people, for instance, because it feels like a further erasure of identity on top of colour-blindness. An important direction for future research is thus to investigate the shift in the gender narrative in different cultural contexts and among participants with multiple, marginalised intersecting identities.

Last, a limitation of the research in this thesis were the methodologies I was able to use. Importantly, due to the covid-19 pandemic, use of the lab was severely limited and recruiting participants for in-person studies was very difficult. While I included diverse methodologies (surveys across all chapters; qualitative observational research in Chapter 2; reaction time tasks in Chapters 3, 4 and 6; eye-tracking in Chapter 6), I was unable to run certain lab studies which I had planned. For instance, the research in Chapter 2 was about social psychological threat, which I had originally wanted to measure using physiological measures of heart rate and blood pressure (Seery, 2013). This is a direction for future research, as it can provide important insights. Specifically, along with measuring threat in an implicit way, one can also use this methodology to measure

challenge: a different physiological state which is often considered to be related to *support* for social change, as opposed to resistance to it (Scheepers & Ellemers, 2019).

Conclusion

In recent years, we have seen a gradual shift in how we conceptualise gender: away from a strict binary, whereby we see everyone as 'men' or 'women', and towards a gender spectrum whereby many identities exist. Through extensive research within this thesis, I have added to our understanding of resistance toward gender-inclusive initiatives, possible positive effects of gender-inclusive initiatives, and how we conceptualise gender for both ourselves and others. These insights further inform us about how we, as a society, can approach the current and future shift in the gender narrative, and inspire follow-up questions for future research.

Footnote

[1] NB: this General Discussion chapter contains all of the elements of a classic scientific dissertation discussion chapter, but may deviate at times from the classic structure of those elements. This chapter is written with practitioners in mind, for instance, D&I officers of organisations, and what they may want to take away from this thesis.

Supplementary Materials

All Supplementary Materials for all chapters can be found via the following link:

https://osf.io/cvbqs/?view_only=48c4eed9c9d64a46864e78a6c4e66b27

Nederlandse Samenvatting

Een vierkant blok in een rond gat duwen: Gender buiten het binaire model

"Wat is je gender, wat voor mens wil je vandaag zijn? Wil je man of vrouw of liever toch neutraal zijn?"

Wat is je gender – Stippenlift

Het hoofddoel van dit proefschrift was om te bekijken wat de gangbare opvattingen zijn over het traditionele genderbinaire model en hoe de samenleving dit systeem geleidelijk aan begint los te laten. Het 'genderbinaire' model is het idee dat mensen in twee categorieën kunnen worden ingedeeld, namelijk mannen en vrouwen, en dat deze categorieën natuurlijk en onveranderlijk zijn (Hyde et al., 2019). Onze samenleving, zowel in Nederland als in andere westerse landen, is gebaseerd op een genderbinaire structuur. Onze officiële identiteitsdocumenten vermelden bijvoorbeeld doorgaans een geslachtsaanduiding die binair is. Uitzonderingen hierop, zoals een X als geslachtsaanduiding, zijn zeldzaam.

Het genderbinaire model weerspiegelt echter niet hoe veel mensen hun eigen genderidentiteit ervaren (Kuyper & Wijsen, 2014; Van Caenegem et al., 2015). Daarom lijkt het gerechtvaardigd dat er een verandering in de samenleving plaatsvindt. Met andere woorden, als veel mensen zich voelen alsof ze het spreekwoordelijke "vierkante blok" zijn dat in een "rond gat" wordt geduwd, is het wellicht verstandig om deze rigide structuren meer los te laten.

In de afgelopen jaren hebben we gezien dat deze verandering steeds meer voet aan de grond krijgt, bijvoorbeeld door een grotere erkenning van diverse genderidentiteiten. Dit komt onder andere tot uiting in de wetgeving (Ryan, 2018), in alledaagse taalgebruik (Bradley et al., 2019; Gustafsson Sendén et al., 2015), in de fysieke omgevingen die we regelmatig bezoeken (Bovens & Marcoci, 2023), en zelfs in de mode (Kim et al., 2022). Deze veranderingen in onze samenleving roepen vragen op over hoe het genderbinaire model heeft standgehouden. Daar bestaan enkele wijdverbreide theoretische veronderstellingen over. Een van die veronderstellingen is dat het genderbinaire model iets 'natuurlijks' is en daardoor moeilijk te veranderen is. Een andere veronderstelling is dat het genderbinaire model handig is om mensen

in sociale categorieën te plaatsen zodat we de sociale wereld om ons heen makkelijker kunnen begrijpen. Deze veronderstellingen worden behandeld in Hoofdstuk 1, de inleiding.

Naast het bespreken van veronderstellingen over hoe het systeem in stand wordt gehouden, weerleg ik ook enkele algemeen aangenomen veronderstellingen met betrekking tot de *verandering* in het gendersysteem. Ik richt me met name op ideeën die professionals zoals beleidsmakers of diversiteitsmedewerkers soms hebben. Bijvoorbeeld, de veronderstelling dat het implementeren van gender-inclusieve initiatieven weinig positieve en veel negatieve effecten zou hebben. Ik betwist deze ideeën in Hoofdstuk 7, de algemene discussie, op basis van de onderzoeksresultaten die in dit proefschrift worden gepresenteerd.

In deze Nederlandse samenvatting richt ik me met name op kernboodschappen voor *professionals*. Ik doe dit door de bevindingen uit elk hoofdstuk te belichten, die vooral voor professionals van belang zijn. Ik zal ook bespreken wat deze bevindingen betekenen voor de praktijk.

Deel 1: Gender-inclusieve initiatieven

Dit proefschrift is opgedeeld in twee delen. In Deel 1 van dit proefschrift, bestaande uit twee hoofdstukken, heb ik me gericht op gender-inclusieve initiatieven. Ik heb gekeken naar discussies die ik hierover tegenkwam in de media en online. Ik heb ook onderzocht welke positieve impact gender-inclusieve initiatieven mogelijk kunnen hebben op de samenleving als geheel.

In Hoofdstuk 2 heb ik online discussies over gender-inclusieve initiatieven onderzocht. Hierbij heb ik gekeken naar zowel het perspectief van degenen die hun meningen hierover in online comments uiten (de 'schrijvers') als degenen die deze online comments mogelijk lezen (de 'lezers'). Een belangrijke bevinding in dit hoofdstuk was dat negatieve meningen over gender-inclusieve

initiatieven mogelijk verspreid worden door een kleine, bedreigde minderheid (de 'schrijvers'). Aan de andere kant zijn de mensen die deze online meningen lezen (de 'lezers') het vaak niet eens met de schrijvers, maar hebben zij ook geen mening die sterk genoeg is om zelf aan de discussie deel te willen nemen. Hierdoor kunnen online discussies over gender-inclusieve initiatieven een vertekend beeld geven van de algemene opinie hierover. In dit hoofdstuk ben ik tot de conclusie gekomen dat bij het beslissen of men een gender-inclusief initiatief wil uitvoeren, het niet verstandig is om te veel te letten op de meningen van mensen die online hun gedachten hierover delen.

In Hoofdstuk 3 heb ik de focus verschoven van de gebruikelijke literatuur die de voordelen van gender-inclusieve initiatieven voor non-binaire en transgender personen benadrukt. Ik heb specifiek onderzocht of dergelijke initiatieven ook verborgen voordelen kunnen hebben voor cisgender mannen en vrouwen. Bijvoorbeeld, door het verminderen van genderstereotypen op de arbeidsmarkt. Het kan namelijk lastig zijn voor mannen om te kiezen voor banen waar vrouwen in de meerderheid zijn, en voor vrouwen om te kiezen voor banen waar mannen in de meerderheid zijn. Dit komt doordat sommige banen nu nog worden gezien als meer geschikt voor het ene geslacht dan voor het andere. Met andere woorden, banen worden gestereotypeerd. In dit hoofdstuk heb ik ontdekt dat gender-inclusieve initiatieven de manier waarop mensen denken over genderstereotypen meteen na implementatie kunnen veranderen. Dit zou goed voor iedereen zijn. Mannen en vrouwen zouden zich daardoor namelijk minder beperkt in hun keuzes voelen en organisaties zouden meer talent kunnen aantrekken, ongeacht gender.

Deel 2: Hoe wij (ons eigen en andermans) gender begrijpen

In Deel 2 van dit proefschrift, bestaande uit drie hoofdstukken, heb ik me gericht op hoe we gender momenteel begrijpen en hoe dit

kan veranderen wanneer onze samenleving minder binair wordt. Hierbij heb ik onderzocht hoe we niet alleen onze eigen gender identificeren, maar ook hoe we de genders van andere mensen begrijpen.

In Hoofdstuk 4 heb ik onderzocht hoe mensen hun eigen gender zien, en of deze zelfbeelden meer in lijn zijn met een binair of niet-binair model van gender. Ook heb ik gekeken naar de impact van een niet-binair zelfbeeld[1] op verschillende aspecten van sociaal welzijn, zoals gevoelens van maatschappelijke inclusie, zelfvertrouwen, levensvoldoening en algemene gemoedstoestand. Een belangrijke bevinding was dat een aanzienlijk aantal deelnemers (soms zelfs rond de helft) zichzelf op een niet-binaire manier zag. Ze identificeerden zich met zowel mannen als vrouwen, of met geen van beiden, of met een derde gender. Deze groep 'niet-binaire' mensen ervoer over het algemeen een lager sociaal welzijn dan de groep die zichzelf op een 'binaire' manier categoriseerde. In dit hoofdstuk heb ik geconcludeerd dat een verandering in het gendersysteem voordelen kan hebben voor veel mensen. Een verandering zou namelijk beter aansluiten bij hun zelfbeeld én positieve gevolgen hebben voor hun sociale welzijn, met name hun gevoelens van maatschappelijke inclusie.

In Hoofdstuk 5 heb ik de bevindingen van Hoofdstuk 4 verder onderzocht door gender-zelfbeeld op een impliciete manier te meten, met behulp van een reactietijdtaak. Impliciete manieren van meten zijn namelijk minder beïnvloedbaar door sociale wenselijkheid. Ook heb ik gekeken naar hoe mensen het gender van anderen impliciet begrijpen. Een belangrijke bevinding was dat ik met Hoofdstuk 5 de resultaten van Hoofdstuk 4 verder kon bevestigen. Bovendien heb ik ontdekt dat de manier waarop mensen anderen in genders onderverdelen ook niet strikt binair is. Dit was een belangrijke aanwijzing dat het veranderen van het gendersysteem wellicht minder moeilijk voor mensen zou zijn dan men zou verwachten. Als mensen namelijk genders van anderen

nu al niet op een strikt binaire manier begrijpen, zouden ze misschien ook niet zoveel moeite hebben met een minder binair systeem.

In Hoofdstuk 6 heb ik diepgaander onderzocht hoe mensen anderen onderverdelen op basis van gender. Ik keek hierbij naar hoe mensen biologische en sociale informatie gebruiken om het gender van een ander te begrijpen. Onder biologische informatie valt bijvoorbeeld iemands gezichtsstructuur en onder sociale informatie valt bijvoorbeeld iemands kleding. Ik heb onderzocht hoe gender-inclusief beleid beïnvloedt hoe mensen biologische en sociale informatie gebruiken als ze het gender van een nieuw persoon moeten begrijpen. Specifiek vroeg ik deelnemers zich voor te stellen dat ze voor een bedrijf werkten met ofwel een genderbinair beleid, waarbij "Meneer" en "Mevrouw" werden gebruikt in communicatie, of een minder binair beleid, waarbij meer genderneutrale taal werd gebruikt. Een belangrijke bevinding was dat mensen vaker sociale informatie gebruikten wanneer ze zich voorstelden dat ze voor het minder binaire bedrijf werkten. Met andere woorden, ze baseerden hun begrip van andermans gender meer op bijvoorbeeld kleding dan op de biologie van iemand. Dit benadrukt hoe gender-inclusiviteit van invloed kan zijn op hoe mensen anderen waarnemen en met hen omgaan, en geeft ons inzicht in wat ons te wachten staat als onze maatschappij verder verandert.

Wat betekent dit voor de praktijk?

Sinds het einde van mijn contract als promovendus ben ik overgestapt van de academische wereld naar een functie als onderzoeker in de praktijk. In mijn nieuwe rol heb ik regelmatig gesprekken met professionals uit diverse vakgebieden, zoals HR-managers, beleidsmakers, medewerkers van overheidsinstanties en diversiteitsmedewerkers. Wanneer ik hen vertel over het onderwerp van mijn proefschrift, merk ik dat sommigen van hen bepaalde aannames hebben waarvan ik denk dat mijn proefschrift wellicht tegenbewijs kan leveren.

Ten eerste uiten sommige professionals bezorgdheid over het implementeren van gender-inclusieve initiatieven uit angst voor mogelijke weerstand binnen en buiten hun organisatie. Hoewel het klopt dat discussies rond deze initiatieven vaak negatief kunnen zijn, kan dit beeld vertekend zijn omdat de discussies grotendeels worden gedomineerd door mensen die sterk tegen gender-inclusieve initiatieven zijn. Verder zijn er echter ook veel mensen die hun opvattingen niet zo uitgesproken uiten, maar wel positief staan tegenover gender-inclusieve initiatieven. Dit laat ik ook in Hoofdstuk 2 van dit proefschrift zien.

Verder uiten sommige professionals dat er niemand in hun organisatie non-binair is en dat er daardoor geen behoefte is aan genderinclusief beleid of initiatieven. Ze gaan ervan uit dat slechts een kleine minderheid van mensen zou profiteren van dergelijke initiatieven en dat de financiële middelen dus beter elders besteed kunnen worden. Het is echter zo dat een derde tot de helft van de bevolking zich waarschijnlijk als niet-binair ziet. Zoals ik aantoon in Hoofdstuk 4, zouden gender-inclusieve initiatieven voordelig kunnen zijn voor het welzijn van deze niet-binaire mensen. Daarnaast zijn er verborgen voordelen van gender-inclusieve initiatieven voor degenen die wél binnen de binaire hokjes passen (cisgender mannen en vrouwen dus). Gender-inclusieve initiatieven kunnen namelijk genderstereotypen bestrijden, bijvoorbeeld over de geschiktheid van banen. Dit laat ik in Hoofdstuk 3 zien.

Ten derde aarzelen praktijkmensen soms om genderinclusief beleid of initiatieven te implementeren omdat ze vrezen dat mensen deze verandering moeilijk zullen vinden. Ze denken dus dat er voor medewerkers veel tijd en moeite in zou zitten om zich aan het nieuwe beleid aan te passen. Zoals aangetoond in Hoofdstukken 4-6, zien mensen gender nu al niet als iets wat volledig binair is. Verder heb ik in Hoofdstuk 6, door het analyseren van reactietijden en gedrag tijdens

het uitvoeren van een taak, aangetoond dat mensen zich best kunnen aanpassen aan een genderinclusief organisatiebeleid. Daarnaast heb ik in Hoofdstuk 3 aanwijzingen gevonden dat gender-inclusieve initiatieven bijna meteen na implementatie de overtuigingen van mensen beïnvloeden. Deze aanpasbaarheid suggereert dat mensen misschien niet zoveel moeite hebben met veranderingen in het gendersysteem.

Al met al zijn veel van de bevindingen in dit proefschrift relevant voor professionals omdat ze gangbare aannames weerleggen en laten zien hoe gender-inclusieve initiatieven een organisatie ten goede kunnen komen. Bovendien wijzen veel van de bevindingen in dit proefschrift erop dat de negatieve effecten van het implementeren van dergelijke initiatieven minder talrijk en ernstig zijn dan professionals wellicht vrezen.

Conclusie

In de afgelopen jaren hebben we een geleidelijke verandering gezien in hoe de samenleving over gender denkt. Hoewel we nog in een maatschappij leven die volgens een genderbinair model is opgesteld, verandert dit steeds meer. Gender-inclusieve initiatieven worden talrijker en er wordt steeds meer over non-binaire genders gesproken. Met het uitgebreide onderzoek binnen dit proefschrift heb ik bijgedragen aan de literatuur over weerstand tegen gender-inclusieve initiatieven, de positieve effecten van gender-inclusieve initiatieven, en hoe we gender conceptualiseren voor zowel onszelf als anderen. Deze inzichten bieden perspectieven over hoe wij, als samenleving, de huidige en toekomstige veranderingen in het gendersysteem kunnen benaderen.

Voetnoot

[1] Een niet-binair zelfbeeld is niet hetzelfde als een non-binaire identiteit. Sommige van de deelnemers die een niet-binair zelfbeeld hadden, identificeerden zich wel expliciet als vrouwen of mannen. Voor meer informatie, zie Hoofdstuk 4.

References

Abed, E. C., Schudson, Z. C., Gunther, O. D., Beischel, W. J., & van Anders, S. M. (2019). Sexual and gender diversity among sexual and gender/sex majorities: Insights via sexual configurations theory. *Archives of Sexual Behavior, 48*(5), 1423-1441. https://doi.org/10.1007/s10508-018-1340-2

Alexander, G. M., & Hines, M. (2002). Sex differences in response to children's toys in nonhuman primates (Cercopithecus aethiops sabaeus). *Evolution and Human Behavior, 23*(6), 467-479. https://doi.org/10.1016/s1090-5138(02)00107-1

Andrews, F. M., & Withey, S. B. (1976). *Social indicators of well-being: Americans' perceptions of life quality.* Springer Science & Business Media. https://doi.org/10.1007/978-1-4684-2253-5

Andrews, N. C. Z., Martin, C. L., Cook, R. E., Field, R. D., & England, D. E. (2019). Exploring dual gender typicality among young adults in the United States. *International Journal of Behavioral Development, 43*(4), 314-321. https://doi.org/10.1177/0165025418811125

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods, 52*(1), 388-407. https://doi.org/10.3758/s13428-019-01237-x

Aparte wc's voor mannen en vrouwen stammen uit de negentiende eeuw. (2020, March 21). *Nu.nl.* https://www.nu.nl/wetenschap/6038882/aparte-wcs-voor-mannen-en-vrouwen-stammen-uit-de-negentiende-eeuw.html

Bailey, D.; Barclay, H.; Barrett, N.; Harty, R. (Executive Producers). (1972-1981). *Russell Harty Plus* [TV series]. London Weekend Television.

Baiocco, R., Antoniucci, C., Basili, E., Pistella, J., Favini, A., Martin, C., & Pastorelli, C. (2022). Perceived Similarity to Gender Groups

Scale: Validation in a Sample of Italian LGB + and Heterosexual Young Adults. *Sexuality Research and Social Policy, 19*(3), 1270-1284. https://doi.org/10.1007/s13178-021-00631-5

Beauvoir, S. D. (2000). One is not born a woman. In K. Oliver, (Ed.), *French Feminism Reader* (pp. 1-34). Lanham, MD: Rowman and Littlefield Publishers.

Bem, S. L. (1974). The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology, 42*(2), 155-162. https://doi.org/10.1037/h0036215

Bem, S. L., & Lewis, S. A. (1975). Sex role adaptability: One consequence of psychological androgyny. *Journal of Personality and Social Psychology, 31*(4), 634-643. https://doi.org/10.1037/h0077098

Bennett, T. (2014). No man's land: Non-binary sex identification in Australian law and policy. *University of New South Wales Law Journal, 37*, 847-873.

Bianco, M. (2015, April 2). Statistics show exactly how many times trans people have attacked you in bathrooms. *MIC*. https://www.mic.com/articles/114066/statistics-show-exactly-how-many-times-trans-people-have-attacked-you-in-bathrooms

Blau, F. D., & Kahn, L. M. (2020). The gender pay gap: Have women gone as far as they can?. *Inequality in the United States: A Reader,* 345-362. https://doi.org/10.4324/9781003071709-31

Block, K., Croft, A., & Schmader, T. (2018). Worth less?: Why men (and women) devalue care-oriented careers. *Frontiers in Psychology, 9*. https://doi.org/10.3389/fpsyg.2018.01353

Blumell, L. E., Huemmer, J., & Sternadori, M. (2019). Protecting the ladies: Benevolent sexism, heteronormativity, and partisanship in online discussions of gender-neutral bathrooms. *Mass Communication and Society, 22*(3), 365-388.

https://doi.org/10.1080/15205436.2018.1547833

Bosson, J. K., Vandello, J. A., & Caswell, T. A. (2013). Precarious manhood. *The SAGE Handbook of Gender and Psychology*, 115-130. https://doi.org/10.4135/9781446269930.n8

Bovens, L., & Marcoci, A. (2023). The gender-neutral bathroom: a new frame and some nudges. *Behavioural Public Policy, 7*(1), 1-24. https://doi.org/10.1017/bpp.2020.23

Bradley, E. D., Salkind, J., Moore, A., & Teitsort, S. (2019). Singular 'they' and novel pronouns: Gender-neutral, nonbinary, or both?. *Proceedings of the Linguistic Society of America, 4*(1), 36. https://doi.org/10.3765/plsa.v4i1.4542

Brook, A. T., Garcia, J., & Fleming, M. A. (2008). The effects of multiple identities on psychological well-being. *Personality and Social Psychology Bulletin, 34*(12), 1588-1600. https://doi.org/10.1177/0146167208324629

Bruce, V., Burton, A. M., Hanna, E., Healey, P., Mason, O., Coombes, A., Fright, R., & Linney, A. (1993). Sex discrimination: how do we tell the difference between male and female faces?. *Perception, 22*(2), 131-152. https://doi.org/10.1068/p220131

Bussey, K., & Bandura, A. (1999). Social cognitive theory of gender development and differentiation. *Psychological Review, 106*(4), 676-713. https://doi.org/10.1037//0033-295x.106.4.676

Butler, J. (2002). *Gender trouble.* Routledge. https://doi.org/10.4324/9780203902752

Campbell, D. T. (1960). Recommendations for APA test standards regarding construct, trait, or discriminant validity. *American Psychologist, 15*(8), 546-553. https://doi.org/10.1037/h0048255

Canessa-Pollard, V., Reby, D., Banerjee, R., Oakhill, J., & Garnham, A. (2022). The development of explicit occupational gender stereotypes in children: Comparing perceived gender ratios and competence beliefs. *Journal of Vocational*

Behavior, 134, 103703. https://doi.org/10.1016/j.jvb.2022.103703

Cartwright, A., Hussey, I., Roche, B., Dunne, J., & Muphy, C. (2017). An investigation into the relationship between the gender binary and occupational discrimination using the implicit relational assessment procedure. *The Psychological Record, 67*(1), 121-130. https://doi.org/10.1007/s40732-016-0212-1

Central Bureau for Statistics. (2019, July 31). Stemgedrag stemgerechtigden *Tweede Kamerverkiezingen 12-09-2012.* Retrieved from https://www.cbs.nl/nl-nl/cijfers/detail/82095NED

Cohen, J. (1992). Quantitative methods in psychology: A power primer. *Psychol. Bull., 112*, 1155-1159.

Colliver, B., Coyle, A., & Silvestri, M. (2019). The 'online othering' of transgender people in relation to 'gender neutral toilets'. *Online Othering,* 215-237. https://doi.org/10.1007/978-3-030-12633-9_9

Cooper, B. (2016). Intersectionality. In Disch, Lisa; Hawkesworth, Mary (eds.). *The Oxford Handbook of Feminist Theory. Oxford University Press.* pp. 385–406.

Cooper, D. (2019). A very binary drama: The conceptual struggle for gender's future. *King's College London Law School Research Paper,* 2019-34.

Corbett, G. (2013). Number of Genders. In: Dryer, Matthew S. & Haspelmath, Martin (eds.) WALS Online (v2020.3) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.7385533 (Available online at http://wals.info/chapter/30, Accessed on 2023-07-19.)

Cuddy, A. J. C., Fiske, S. T., & Glick, P. (2008). Warmth and Competence as Universal Dimensions of Social Perception: The Stereotype Content Model and the BIAS Map. *Advances in Experimental Social Psychology*, 61-149. https://doi.org/10.1016/s0065-2601(07)00002-0

Dardenne, B., Dumont, M., & Bollier, T. (2007). Insidious dangers of benevolent sexism: Consequences for women's performance.

Journal of Personality and Social Psychology, 93(5), 764 -779. https://doi.org/10.1037/0022-3514.93.5.764

Davis, E. (2023, June 5). How to use gender-inclusive language in the workplace. *The HR Digest*. https://www.thehrdigest.com/how-to-use-gender-inclusive-language-in-the- workplace/

De Kraker, J., Kuijs, S., Cörvers, R., & Offermans, A. (2014). Internet public opinion on climate change: A world views analysis of online reader comments. *International Journal of Climate Change Strategies and Management, 6*(1), 19-33. https://doi.org/10.1108/ijccsm-09-2013-0109

Dean, M. L., & Tate, C. C. (2017). Extending the legacy of Sandra Bem: Psychological androgyny as a touchstone conceptual advance for the study of gender in psychological science. *Sex Roles, 76* (11-12), 643-654. https://doi.org/10.1007/s11199-016-0713-z

Derks, B., Van Laar, C., Ellemers, N., & De Groot, K. (2011). Gender-bias primes elicit queen-bee responses among senior policewomen. *Psychological Science, 22*(10), 1243-1249. https://doi.org/10.1177/0956797611417258

Dhejne, C., Van Vlerken, R., Heylens, G., & Arcelus, J. (2016). Mental health and gender dysphoria: A review of the literature. *International Review of Psychiatry, 28*(1), 44-57. https://doi.org/10.3109/09540261.2015.1115753

Diener, E. D., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. J*ournal of Personality Assessment, 49*(1), 71-75. https://doi.org/10.1207/s15327752jpa4901_13

Domen, I., Derks, B., Van Veelen, R., & Scheepers, D. (2020). Gender identity relevance predicts preferential neural processing of same-gendered faces. *Social Neuroscience, 15*(3), 334-347. https://doi.org/10.1080/17470919.2019.1703807

Duffy, J., Warren, K., & Walsh, M. (2001). Classroom interactions: Gender of teacher, gender of student, and classroom subject.

Sex roles, 45(9), 579-593.

Duggan, J. (2022). Transformative readings: Harry Potter fan fiction, trans/queer reader response, and J. K. Rowling. *Children's Literature in Education, 53*(2), 147-168. https://doi.org/10.1007/s10583-021-09446-9

Dunne, P., & Mulder, J. (2018). Beyond the binary: Towards a third sex category in Germany? *German Law Journal, 19*(3), 627-648. https://doi.org/10.1017/s2071832200022811

Eagly, A. H. (1983). Gender and social influence: A social psychological analysis. *American Psychologist*, 38(9), 971-981. https://doi.org/10.1037//0003-066x.38.9.971

Ellemers, N., Kortekaas, P., & Ouwerkerk, J. W. (1999). Self-categorisation, commitment to the group and group self-esteem as related but distinct aspects of social identity. *European Journal of Social Psychology, 29*(23), 371-389. https://doi.org/10.1002/(sici)1099-0992(199903/05)29:2/3<371::aid-ejsp932>3.3.co;2-l

Ellemers, N., Spears, R., & Doosje, B. (1999). Social identity: *Context, commitment, content. Blackwell: Oxford.*

Endendijk, J. J., Andrews, N. C., England, D. E., & Martin, C. L. (2019). Gender-identity typologies are related to gender-typing, friendships, and social-emotional adjustment in Dutch emerging adults. *International Journal of Behavioral Development, 43*(4), 322-333. https://doi.org/10.1177/0165025418820686

England, P. (2010). The gender revolution: Uneven and stalled. *Gender & Society, 24*(2), 149-166. https://doi.org/10.1177/0891243210361475

Enke, A. F. (2016). Stick Figures and Little Bits. In Y. Martinez-San Miguel & S. Tobias (Eds.). *Trans studies: The challenge to hetero/homo normativities.* Rutgers University Press.

Estrada, G. S. (2011). Two spirits, Nadleeh, and LGBTQ2 Navajo gaze. *American Indian Culture and Research Journal, 35*(4), 167-190.

https://doi.org/10.17953/aicr.35.4.x500172017344j30

Faniko, K., Ellemers, N., Derks, B., & Lorenzi-Cioldi, F. (2017). Nothing changes, really: Why women who break through the glass ceiling end up reinforcing it. *Personality and Social Psychology Bulletin, 43*(5), 638-651. https://doi.org/10.1177/0146167217695551

Fine, C. (2017). Testosterone Rex: myths of sex, science, and society. New York, NY: WW Norton & Company.

Fiske, S. T., & Macrae, C. N. (Eds.). (2012). *The SAGE handbook of social cognition*. Sage. https://doi.org/10.4135/9781446247631

Fiske, S. T., Cuddy, A. J., Glick, P., & Xu, J. (2002). Competence and warmth scales. *Journal of Personality and Social Psychology.* https://doi.org/10.1037/t35954-000

Francis, J., Sachan, P., Waters, Z., Trapp, G., Pearce, N., Burns, S., Lin, A., & Cross, D. (2022). Gender-neutral toilets: A qualitative exploration of inclusive school environments for sexuality and gender diverse youth in Western Australia. *International Journal of Environmental Research and Public Health, 19*(16), 10089. https://doi.org/10.3390/ijerph191610089

Freeman, J. B., Rule, N. O., Adams, R. B., Jr., & Ambady, N. (2010). The neural basis of categorical face perception: Graded representations of face gender in fusiform and orbitofrontal cortices. *Cerebral Cortex, 20*(6), 1314-1322. https://doi.org/10.1093/cercor/bhp195

Fusar-Poli, P., Placentino, A., Carletti, F., Allen, P., Landi, P., Abbamonte, M., Barale, F., Perez, J., McGuire, P., & Politi, P. L. (2009). Laterality effect on emotional faces processing: ALE meta-analysis of evidence. *Neuroscience Letters, 452*(3), 262-267. https://doi.org/10.1016/j.neulet.2009.01.065

Fusion's Massive Millennial Survey (2015) *Fusion.net.* Retrieved from http://fusion.net/event/massive-millennial-poll/

Garelick, A. S., Filip-Crawford, G., Varley, A. H., Nagoshi, C. T., Nagoshi, J. L.,

& Evans, R. (2017). Beyond the binary: Exploring the role of ambiguity in biphobia and transphobia. *Journal of Bisexuality, 17*, 1-18. https://doi.org/10.1080/15299716.2017.1319890

Garfinkel, H. (1967). Passing and the managed achievement of sex status in an "intersexed person". IN: ders. *Studies in Ethnomethodology (S. 116–185)*. Cambridge, Polity Press.

Garrett, C. S., Ein, P. L., & Tremaine, L. (1977). The development of gender stereotyping of adult occupations in elementary school children. *Child Development , 48*(2), 507. https://doi.org/10.2307/1128646

Gawronski, B., & Payne, B. K. (Eds.). (2010). Handbook of implicit social cognition: Measurement, theory, and applications. New York, NY: Guilford Press.

Gawronski, B., & Sritharan, R. (2010). Formation, change, and contextualization of mental associations: Determinants and principles of variations in implicit measures. In B. Gawronski & B. K. Payne (Eds.), Handbook of implicit social cognition: Measurement, theory, and applications (pp. 216-240). New York, NY: Guilford Press.

Gawronski, B., Morrison, M., Phills, C. E., & Galdi, S. (2017). Temporal stability of implicit and explicit measures: A longitudinal analysis. *Personality and Social Psychology Bulletin, 43*(3), 300-312. https://doi.org/10.1177/0146167216684131

Gershenson, O. (2010). The restroom revolution: Unisex toilets and campus politics. In H. Molotch & L. Norén (Eds.), *Toilet: Public restrooms and the politics of sharing* (Vol. 1, pp. 191-207). New York University Press.

Gibson, S., & Fernandez, J. (2018). *Gender diversity and non-binary inclusion in the workplace: The essential guide for employers.* Jessica Kingsley Publishers.

Glick, P., & Fiske, S. T. (1996). The ambivalent sexism inventory: Differentiating hostile and benevolent sexism. *Journal*

of Personality and Social Psychology, 70(3), 491-512. https://doi.org/10.1037//0022-3514.70.3.491

Goldman, B. M., & Kernis, M. H. (2002). The role of authenticity in healthy psychological functioning and subjective well-being. *Annals of the American Psychotherapy Association,* 5(6), 18-20.

Goodman, E., Adler, N. E., Kawachi, I., Frazier, A. L., Huang, B., & Colditz, G. A. (2001). Adolescents' perceptions of social status: development and evaluation of a new indicator. *Pediatrics, 108*(2), e31-e31. https://doi.org/10.1542/peds.108.2.e31

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the implicit association test: I. An improved scoring algorithm. *Journal of personality and social psychology, 85*(2), 197-216. https://doi.org/10.1037/0022-3514.85.2.197

Gregg, A. P., Seibt, B., & Banaji, M. R. (2006). Easier done than undone: asymmetry in the malleability of implicit preferences. *Journal of personality and social psychology, 90*(1), 1-20. https://doi.org/10.1037/0022-3514.90.1.1

Gustafsson Sendén, M., Bäck, E. A., & Lindqvist, A. (2015). Introducing a gender-neutral pronoun in a natural gender language: the influence of time on attitudes and behavior. *Frontiers in psychology,* 6. https://doi.org/10.3389/fpsyg.2015.00893

Haines, E. L., & Stroessner, S. J. (2019). The role prioritization model: How communal men and agentic women can (sometimes) have it all. *Social and Personality Psychology Compass, 13*(12), e12504. https://doi.org/10.1111/spc3.12504

Hawkins, E. (2022, July 20). Argos shoppers and staff accuse retailer of 'pink washing' over lack of gender-neutral 'Mx' option. *City A.M..* https://www.cityam.com/sainsburys-asks-argos-customers-and-staff-to-bear-with-us-amid-criticism-for-no-gender-neutral-options/

Hayes, T. R., & Reiman, A. K. (2022). Evoking gender distinctiveness

threat in cisgender women lowers their support for gender-inclusive bathroom policies. *Group Processes & Intergroup Relations, 25*(7), 1760-1780. https://doi.org/10.1177/13684302211042417

Hord, L. C. (2016). Bucking the linguistic binary: Gender neutral language in English, Swedish, French, and German. *Western Papers in Linguistics/Cahiers linguistiques de Western, 3*(1), 4.

Huart, J., Corneille, O., & Becquart, E. (2005). Face-based categorization, context-based categorization, and distortions in the recollection of gender ambiguous faces. J*ournal of Experimental Social Psychology, 41*(6), 598-608. https://doi.org/10.1016/j.jesp.2004.10.007

Hunt, S. (2018). Embodying self-determination: Beyond the gender binary. *Determinants of Indigenous Peoples' Health, 22-39.*

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. https://doi.org/10.1037/amp0000307

Ito, T. A., & Urland, G. R. (2003). Race and gender on the brain: electrocortical measures of attention to the race and gender of multiply categorizable individuals. *Journal of personality and social psychology, 85*(4), 616-626. https://doi.org/10.1037/0022-3514.85.4.616

J.K. Rowling nuanceert uitspraken: 'Transgenders moeten beschermd worden'. (2020, June 11). *Nu.nl.* https://www.nu.nl/achterklap/6057206/jk-rowling-nuanceert-uitspraken-transgenders-moeten-beschermd-worden.html

Jackson, M. C., Wilde, V. K., & Goff, P. A. (2016). Seeing color blindness: Color-blind racial ideology research methods in social psychology. *The myth of racial color blindness: Manifestations, dynamics, and impact.,* 125-140.

https://doi.org/10.1037/14754-008

Jansen, W. S., Otten, S., Van Der Zee, K. I., & Jans, L. (2014). Inclusion: Conceptualization and measurement. *European Journal of Social Psychology, 44*(4), 370-385. https://doi.org/10.1002/ejsp.2011

Jaroszewski, S., Lottridge, D., Haimson, O. L., & Quehl, K. (2018, April). "Genderfluid" or" Attack Helicopter" Responsible HCI Research Practice with Non-binary Gender Variation in Online Communities. In P*roceedings of the 2018 CHI conference on human factors in computing systems* (pp. 1-15).

Jeffreys, S. (2014). *Gender hurts: A feminist analysis of the politics of transgenderism.* Routledge.

Jetten, J., Haslam, S. A., & Haslam, C. (2012). The case for a social identity analysis of health and well-being. In *The social cure: Identity health and well-being* (1st ed., pp. 3-19). Psychology Press.

Joel, D., Tarrasch, R., Berman, Z., Mukamel, M., & Ziv, E. (2014). Queering gender: Studying gender identity in 'normative' individuals. *Psychology & Sexuality, 5*(4), 291-321. https://doi.org/10.1080/19419899.2013.830640

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), 498-509. https://doi.org/10.1037/0022-3514.88.3.498

Jost, J. T., & Van der Toorn, J. (2012). System justification theory. *Handbook of theories of social psychology*, 313-343. https://doi.org/10.4135/9781446249222.n42

Jost, J. T., Liviatan, I., van der Toorn, J., Ledgerwood, A., Mandisodza, A., & Nosek, B. A. (2010). System justification: How do we know it's motivated?. In Bobocel, D. R., Kay, A. C., Zanna, M. P., & Olson, J. M. (Eds.), *The psychology of justice and legitimacy*. Psychology Press.

Kasai, M. (2023). Inter-minority empathy. In Kasai, M., Toda, Y., Russell, S. (Eds.) *SOGI minority and school life in Asian contexts* (pp. 198-208). Routledge. https://doi.org/10.4324/9781003430414-12

Kecskés-Kovács, K., Sulykos, I., & Czigler, I. (2013). Is it a face of a woman or a man? Visual mismatch negativity is sensitive to gender category. *Frontiers in Human Neuroscience*, 7, 532. https://doi.org/10.3389/fnhum.2013.00532

Keohane, A., & Richardson, N. (2018). Negotiating gender norms to support men in psychological distress. *American journal of men's health*, 12(1), 160-171. https://doi.org/10.1177/1557988317733093

Kessler, S. J., & McKenna, W. (1985). *Gender: An ethnomethodological approach.* University of Chicago Press.

Kim, H., Cho, I., & Park, M. (2022). Analyzing genderless fashion trends of consumers' perceptions on social media: Using unstructured big data analysis through Latent Dirichlet Allocation-based topic modeling. *Fashion and Textiles*, 9(1), 1-21. https://doi.org/10.1186/s40691-021-00281-6

Klasen, S. (2007). Gender-related indicators of well-being. In *Human well-being*: Concept and measurement (pp. 167-192). London: Palgrave Macmillan UK. https://doi.org/10.1057/9780230625600_7

Kristof-Brown, A. L., Zimmerman, R. D., & Johnson, E. C. (2005). Consequences of Individuals' Fit at Work: a Meta-analysis of Person–Job, Person–Organization, Person–Group, and Person–Supervisor Fit. *Personnel Psychology, 58*(2), 281-342. https://doi.org/10.1111/j.1744-6570.2005.00672.x

Krueger, J. I., & DiDonato, T. E. (2008). Social categorization and the perception of groups and group differences. S*ocial and Personality Psychology Compass, 2*(2), 733-750. https://doi.org/10.1111/j.1751-9004.2008.00083.x

Kuyper, L., & Wijsen, C. (2014). Gender identities and gender dysphoria in the Netherlands. *Archives of Sexual Behavior, 43*(2), 377-385. https://doi.org/10.1007/s10508-013-0140-y

Lapidot-Lefler, N., & Barak, A. (2012). Effects of anonymity, invisibility, and lack of eye-contact on toxic online disinhibition. *Computers in Human Behavior, 2*8(2), 434-443. https://doi.org/10.1016/j.chb.2011.10.014

Latinus, M., VanRullen, R., & Taylor, M. J. (2010). Top-down and bottom-up modulation in processing bimodal face/voice stimuli. *BMC Neuroscience, 11*(1). https://doi.org/10.1186/1471-2202-11-36

Leach, C. W., Van Zomeren, M., Zebel, S., Vliek, M. L., Pennekamp, S. F., Doosje, B., Ouwerkerk, J. W., & Spears, R. (2008). Group-level self-definition and self-investment: A hierarchical (multicomponent) model of in-group identification. *Journal of Personality and Social Psychology, 95*(1), 144-165. https://doi.org/10.1037/0022-3514.95.1.144

Lee, S., Atkinson, L., & Sung, Y. H. (2022). Online bandwagon effects: Quantitative versus qualitative cues in online comments sections. *New Media & Society, 24*(3), 580-599. https://doi.org/10.1177/1461444820965187

Leinbach, M. D., & Fagot, B. I. (1993). Categorical habituation to male and female faces: Gender schematic processing in infancy. *Infant Behavior and Development, 16*(3), 317-332. https://doi.org/10.1016/0163-6383(93)80038-a

Lemmen, I. C., van. (2021). On Gender-Inclusive Language and Its Opponents' Motivations [Master thesis, Utrecht University]. Utrecht University Student Theses Repository. https://studenttheses.uu.nl/handle/20.500.12932/40475

Lenton, A. P., Bruder, M., & Sedikides, C. (2009). A meta-analysis on the malleability of automatic gender stereotypes. *Psychology of Women Quarterly, 33*(2), 183-196.

https://doi.org/10.1111/j.1471-6402.2009.01488.x

Levi, J. L. (2006). Clothes don't make the man (or woman), but gender identity might. *Colum. J. Gender & L., 15*, 90.

Lombardi, E. (2009). Varieties of transgender/transsexual lives and their relationship with transphobia. *Journal of Homosexuality, 56*(8), 977-992. https://doi.org/10.1080/00918360903275393

Luo, G., Houston, V. L., Mussman, M., Garbarini, M., Beattie, A. C., & Thongpop, C. (2009). Comparison of male and female foot shape. *Journal of the American Podiatric Medical Association, 99*(5), 383-390. https://doi.org/10.7547/0990383

Malinowska, A. (2020). Waves of feminism. *The International Encyclopedia of Gender, Media, and Communication,* 1, 1-7.

Mallett, R. K., Huntsinger, J. R., Sinclair, S., & Swim, J. K. (2008). Seeing through their eyes: When majority group members take collective action on behalf of an outgroup. *Group Processes & Intergroup Relations, 11*(4), 451-470. https://doi.org/10.1177/1368430208095400

Mallett, R. K., Huntsinger, J. R., & Swim, J. K. (2011). The role of system-justification motivation, group status and system threat in directing support for hate crimes legislation*. Journal of Experimental Social Psychology, 4*7(2), 384-390. https://doi.org/10.1016/j.jesp.2010.10.014

Marrewijk, T., & Einder, K. (2021, May 19). Raven van Dorst is non-binair: wat betekent dat eigenlijk? *Nu.nl.* https://www.nu.nl/gezondheid/6134356/raven-van-dorst-is-non-binair-wat-betekent-dat-eigenlijk.html

Martin, C. L., & Halverson, C. F. (1983). Gender constancy: A methodological and theoretical analysis. *Sex Roles, 9*, 775-790. https://doi.org/10.1007/bf00290030

Martin, C. L., & Halverson, C. F., Jr. (1983). The effects of sex-typing schemas on young children's memory. *Child Development,*

54(3), 563. https://doi.org/10.2307/1130043

Martin, C. L., Andrews, N. C., England, D. E., Zosuls, K., & Ruble, D. N. (2017). A dual identity approach for conceptualizing and measuring children's gender identity. *Child Development, 88*(1), 167-182. https://doi.org/10.1111/cdev.12568

McGann, H. (2017, July 13). London Tube to welcome passengers with gender-neutral announcements. *CNN.* https://edition.cnn.com/2017/07/13/health/london-underground-tube-gender-neutral/index.html

McKinnon, R. (2018). The epistemology of propaganda. *Philosophy and Phenomenological Research, 96*(2), 483-489. https://doi.org/10.1111/phpr.12429

Merriam-Webster. (n.d.). Mx. – *A gender-neutral honorific.* https://www.merriam-webster.com/wordplay/mx-gender-neutral-title

Morgenroth, T., & Ryan, M. K. (2018). Gender trouble in social psychology: How can Butler's work inform experimental social psychologists' conceptualization of gender?. *Frontiers in Psychology, 9*. https://doi.org/10.3389/fpsyg.2018.01320

Morgenroth, T., & Ryan, M. K. (2021). The effects of gender trouble: An integrative theoretical framework of the perpetuation and disruption of the gender/sex binary. *Perspectives on Psychological Science, 16*(6), 1113-1142. https://doi.org/10.1177/1745691620902442

Morgenroth, T., Sendén, M. G., Lindqvist, A., Renström, E. A., Ryan, M. K., & Morton, T. A. (2021). Defending the sex/gender binary: The role of gender identification and need for closure. *Social Psychological and Personality Science, 12*(5), 731-740. https://doi.org/10.1177/1948550620937188

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), 140-151. https://doi.org/10.1037/a0018093

Mouchetant-Rostaing, Y., Giard, M. H., Bentin, S., Aguera, P. E., & Pernier, J. (2000). Neurophysiological correlates of face gender processing in humans. *European Journal of Neuroscience, 12*(1), 303-310. https://doi.org/10.1046/j.1460-9568.2000.00888.x

Nadal, K. L., Skolnik, A., & Wong, Y. (2012). Interpersonal and systemic microaggressions toward transgender people: Implications for counseling. *Journal of LGBT Issues in Counseling, 6*(1), 55-82. https://doi.org/10.1080/15538605.2012.648583

Nelson, J. A., Liss, M., Erchull, M. J., Hurt, M. M., Ramsey, L. R., Turner, D. L., & Haines, M. E. (2008). Identity in action: Predictors of feminist self-identification and collective action. *Sex Roles, 58*, 721-728. https://doi.org/10.1007/s11199-007-9384-0

Newman, J. (1995). Gender and cultural change. In *Gender, culture and organizational change* (pp. 10-29). Routledge. https://doi.org/10.4324/9780203427965_chapter_1

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological review, 84*(3), 231-259. https://doi.org/10.1037//0033-295x.84.3.231

NOM bereiksonderzoek en oplageregistratie dagbladen magazines en vaktijdschriften totaal lezen media. (n.d.). https://www.nommedia.nl/

Nye, B. (2017). *The sexual spectrum. Netflix.* https://www.netflix.com/title/80117748

OECD (2021), Gender wage gap (indicator). doi: 10.1787/7cee77aa-en (Accessed on 20 July 2021) https://data.oecd.org/earnwage/gender-wage-gap.htm

Outten, H. R., Lee, T., & Lawrence, M. E. (2019). Heterosexual women's support for trans-inclusive bathroom legislation depends on the degree to which they perceive trans women as a threat.

Group Processes & Intergroup Relations, 22(8), 1094-1108. https://doi.org/10.1177/1368430218812660

Palan, S., & Schitter, C. (2018). Prolific.ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance, 17,* 22-27. https://doi.org/10.1016/j.jbef.2017.12.004

Papoutsaki, A., Sangkloy, P., Laskey, J., Daskalova, N., Huang, J., & Hays, J. (2016). WebGazer. Js: Democratizing webcam eye tracking on the browser. *WebGazer. Js.*

Park, J. (2014). *Unisex Clothing*. Love to Know. https://fashion-history.lovetoknow.com/clothing-types-styles/unisex-clothing

Pauletti, R. E., Menon, M., Cooper, P. J., Aults, C. D., & Perry, D. G. (2017). Psychological androgyny and children's mental health: A new look with new measures. *Sex Roles, 76* (11-12), 705-718. https://doi.org/10.1007/s11199-016-0627-9

Paulhus, D. L. (2017). Socially desirable responding on self-reports. *Encyclopedia of Personality and Individual Differences,* 1-5. https://doi.org/10.1007/978-3-319-28099-8_1349-1

Payne, B. K., Burkley, M. A., & Stokes, M. B. (2008). Why do implicit and explicit attitude tests diverge? The role of structural fit. *Journal of personality and social psychology, 94*(1), 16-31. https://doi.org/10.1037/0022-3514.94.1.16

Pearce, R., Erikainen, S., & Vincent, B. (2020). TERF wars: An introduction. *The Sociological Review, 68*(4), 677-698. https://doi.org/10.1177/0038026120934713

Pjesivac, I., Geidner, N., & Cameron, J. (2018). Social credibility online: The role of online comments in assessing news article credibility. *Newspaper Research Journal, 39*(1), 18-31. https://doi.org/10.1177/0739532918761065

Porten-Cheé, P., & Eilders, C. (2015). Spiral of silence online: How online communication affects opinion climate perception and opinion expression regarding the climate change

debate. *Studies in Communication Sciences, 15*(1), 143-150. https://doi.org/10.1016/j.scoms.2015.03.002

Postmes, T., Haslam, S. A., & Jans, L. (2013). A single-item measure of social identification: Reliability, validity, and utility. *British journal of social psychology, 52*(4), 597-617. https://doi.org/10.1111/bjso.12006

Pronin, E. (2008). How we see ourselves and how we see others. *Science, 320*(5880), 1177-1180. https://doi.org/10.1126/science.1154199

Pronin, E., Gilovich, T., & Ross, L. (2004). Objectivity in the eye of the beholder: divergent perceptions of bias in self versus others. *Psychological Review, 111*(3), 781-799. https://doi.org/10.1037/0033-295x.111.3.781

Prusaczyk, E., & Hodson, G. (2020). The roles of political conservatism and binary gender beliefs in predicting prejudices toward gay men and people who are transgender. *Sex Roles, 82*, 438-446. https://doi.org/10.1007/s11199-019-01069-1

Pun, R., Flash, K. S., Nangle, D., & Hernandez Perez, J. (2017). Libraries and Gender Neutral/Inclusive Bathrooms: Case Studies on Promoting Information, Inclusivity and Access in Open Spaces.

Raymond, J. G. (1980). *The transsexual empire*. London: Women's Press.

Reedy, A. (2020, January 30). Sex determination: More complicated than you thought. Ted-Ed. https://www.ted.com/talks/aaron_reedy_sex_determination_more_complicated_than_you_thought

Reis, B., Pereira, M., Azevedo, S., Jerónimo, N., & Miguel, R. (2018). Genderless clothing issues in fashion. In *Textiles, Identity and Innovation: Design the Future* (pp. 255-260). CRC Press. https://doi.org/10.1201/9781315100210-47

Richard, F. D., Bond, C. F. Jr., & Stokes-Zoota, J. J. (2003). One hundred years of social psychology quantitatively described. *Review of General Psychology, 7*(4), 331-363. https://doi.org/10.1037/1089-2680.7.4.331

Richards, C., Bouman, W. P., & Barker, M. (2017). Non-binary genders. *London: Pal grave Macmillan.*

Richards, C., Bouman, W. P., & Barker, M.-J. (Eds.). (2017). Genderqueer and Non-Binary Genders. Palgrave Macmillan UK. https://doi.org/10.1057/978-1-137-51053-2

Rosenberg, M. (1979). Conceiving the Self. Basic Books.

Rule, N. O., & Sutherland, S. L. (2017). Social categorization from faces: Evidence from obvious and ambiguous groups. *Current Directions in Psychological Science, 26*(3), 231-236. https://doi.org/10.1177/0963721417697970

Ryan, J. (2018). Gender identity laws: The legal status of global sex/gender identity recognition. *LGBTQ Policy Journal, 8*, 3-16.

Saguy, T., Reifen-Tagar, M., & Joel, D. (2021). The gender-binary cycle: the perpetual relations between a biological-essentialist view of gender, gender ideology, and gender-labelling and sorting. *Philosophical Transactions of the Royal Society B, 376*(1822), 20200141. https://doi.org/10.1098/rstb.2020.0141

Şahin, O., Van Der Toorn, J., Jansen, W. S., Boezeman, E. J., & Ellemers, N. (2019). Looking beyond our similarities: How perceived (in) visible dissimilarity relates to feelings of inclusion at work. *Frontiers in Psychology, 10.* https://doi.org/10.3389/fpsyg.2019.00575

Sawyer, K., Thoroughgood, C., & Webster, J. (2016). Queering the gender binary: Understanding transgender workplace experiences. *Sexual Orientation and Transgender Issues in Organizations*, 21-42. https://doi.org/10.1007/978-3-319-29623-4_2

Scheepers, D., & Ellemers, N. (2019). Status stress: Explaining defensiveness to the resolution of social inequality in members of dominant groups. *The Social Psychology of Inequality*, 267-287. https://doi.org/10.1007/978-3-030-28856-3_17

Scheepers, D., Ellemers, N., & Sintemaartensdijk, N. (2009). Suffering from

the possibility of status loss: Physiological responses to social identity threat in high status groups. *European Journal of Social Psychology, 39*(6), 1075-1092. https://doi.org/10.1002/ejsp.609

Schubert, T. W., & Otten, S. (2002). Overlap of self, ingroup, and outgroup: Pictorial measures of self-categorization. *Self and Identity, 1*(4), 353-376. *https://doi.org/10.1080/152988602760328012*

Schudson, Z. C., Beischel, W. J., & Van Anders, S. M. (2019). Individual variation in gender/sex category definitions. *Psychology of Sexual Orientation and Gender Diversity, 6*(4), 448-460. https://doi.org/10.1037/sgd0000346

Seery, M. D. (2013). The biopsychosocial model of challenge and threat: Using the heart to measure the mind. S*ocial and Personality Psychology Compass, 7*(9), 637-653. https://doi.org/10.1111/spc3.12052

Segato, R. L., & Monque, P. (2021). Gender and coloniality: From low-intensity communal patriarchy to high-intensity colonial-modern patriarchy. *Hypatia, 36*(4), 781-799. https://doi.org/10.1017/hyp.2021.58

Sendén, M. G., Renström, E., & Lindqvist, A. (2021). Pronouns beyond the binary: The change of attitudes and use over time. *Gender & Society, 35*(4), 588-615. https://doi.org/10.1177/08912432211029226

Sendén, M., Bäck, E. A., & Lindqvist, A. (2015). Introducing a gender-neutral pronoun in a natural gender language: the influence of time on attitudes and behavior. *Frontiers in Psychology, 6*, 893. https://doi.org/10.3389/fpsyg.2015.00893

Simon, B. (1992). The perception of ingroup and outgroup homogeneity: Reintroducing the intergroup context. *European Review of Social Psychology, 3*(1), 1-30. https://doi.org/10.1080/14792779243000005

Singhal, A., & Rogers, E. (2012). *Entertainment-education*: A communication strategy for social change. Routledge. https://doi.org/10.4324/9781410607119

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. https://doi.org/10.1371/journal.pone.0200921

Skinner, G. (2018). The perils of perception 2018. Ipsos. https://www.ipsos.com/en-uk/perils-perception-2018

Smeding, A. (2012). Women in science, technology, engineering, and mathematics (STEM): An investigation of their implicit gender stereotypes and stereotypes' connectedness to math performance. *Sex Roles, 67*(11-12), 617-629. https://doi.org/10.1007/s11199-012-0209-4

Smith, M. (2020, July 16). Where does the British public stand on transgender rights? YouGov. https://yougov.co.uk/topics/politics/articles-reports/2020/07/16/where-does-british-public-stand-transgender-rights

Somai, R. S., Schut, M. J., & Van der Stigchel, S. (2020). Evidence for the world as an external memory: A trade-off between internal and external visual memory storage. *Cortex, 122*, 108-114. https://doi.org/10.1016/j.cortex.2018.12.017

Spears, R., Doosje, B., & Ellemers, N. (1997). Self-stereotyping in the face of threats to group status and distinctiveness: The role of group identification. *Personality and Social Psychology Bulletin, 23*(5), 538-553. https://doi.org/10.1177/0146167297235009

Stern, C., & Rule, N. O. (2018). Physical androgyny and categorization difficulty shape political conservatives' attitudes toward transgender people. *Social Psychological and Personality Science, 9*(1), 24-31. https://doi.org/10.1177/1948550617703172

Steyer, R., Schmitt, M., & Eid, M. (1999). Latent state–trait theory and research in personality and individual differences. European Journal of Personality, 13(5), 389-408. https://doi.org/10.1002/(sici)1099-0984(199909/10)13:5<389::aid-per361>3.0.co;2-a

Stock, K. (2021). *Material girls: Why reality matters for feminism.* Hachette UK.

Stryker, S. (2008). *Transgender history*. Seal Press.

Sutton, A. (2020). Living the good life: A meta-analysis of authenticity, well-being and engagement. *Personality and Individual Differences, 153*, 109645. https://doi.org/10.1016/j.paid.2019.109645

Swerling, G. (2023, February 7). Gender-neutral God to be considered by Church of England. *The Telegraph.* https://www.telegraph.co.uk/news/2023/02/07/church-england-consider-use-gender-neutral-language/

Tajfel, H. (1974). Social identity and intergroup behaviour. *Social Science Information, 13*(2), 65-93. https://doi.org/10.1177/053901847401300204

Tellhed, U., Bäckström, M., & Björklund, F. (2017). Will I fit in and do well? The importance of social belongingness and self-efficacy for explaining gender differences in interest in STEM and HEED majors. *Sex Roles, 77*(1-2), 86-96. https://doi.org/10.1007/s11199-016-0694-y

Torre, M. (2018). Stopgappers? The occupational trajectories of men in female-dominated occupations. *Work and Occupations, 45*(3), 283-312. https://doi.org/10.1177/0730888418780433

Tsouroufli, M. (2002). Gender and teachers' classroom practice in a secondary school in Greece. *Gender and education, 14*(2), 135-147. https://doi.org/10.1080/09540250220133996

Turnbull-Dugarte, S. J., & McMillan, F. (2023). "Protect the women!" Trans-exclusionary feminist issue framing and support for

transgender rights. *Policy Studies Journal, 51*(3), 629-666. https://doi.org/10.1111/psj.12484

Turner, J. C., & Reynolds, K. J. (2011). Self-categorization theory. In P. A. Van Lange, E. T. Higgins & A. W. Kruglanski (Eds.). *Handbook of theories of social psychology: Geographical perspectives* (pp. 399-417). Sage Publications Ltd. https://doi.org/10.4135/9781446249222.n46

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering the social group: A self-categorization theory*. Oxford, England: Blackwell.

Uhlmann, E. L., Leavitt, K., Menges, J. I., Koopman, J., Howe, M., & Johnson, R. E. (2012). Getting explicit about the implicit: A taxonomy of implicit measures and guide for their use in organizational research. *Organizational Research Methods, 15*(4), 553-601. https://doi.org/10.1177/1094428112442750

Ura, A. (2015, November 3). Bathroom fears flush Houston discrimination ordinance. *The Texas Tribune.* https://www.texastribune.org/2015/11/03/houston-anti-discrimination-ordinance-early-voting/

Van Bavel, J. J., & Cunningham, W. A. (2010). A social neuroscience approach to self and social categorisation: A new look at an old issue. *European Review of Social Psychology, 21*(1), 237-284. https://doi.org/10.1080/10463283.2010.543314

Van Breen, J. A., Spears, R., Kuppens, T., & de Lemus, S. (2017). A multiple identity approach to gender: Identification with women, identification with feminists, and their interaction. *Frontiers in Psychology, 8*, 1019. https://doi.org/10.3389/fpsyg.2017.01019

Van Caenegem, E., Wierckx, K., Elaut, E., Buysse, A., Dewaele, A., Van Nieuwerburgh, F., De Cuypere, G., & T'Sjoen, G. (2015). Prevalence of gender nonconformity in Flanders, Belgium. *Archives of Sexual Behavior, 44*(5), 1281-1287.

https://doi.org/10.1007/s10508-014-0452-6

Van Nunspeet, F., Ellemers, N., Derks, B., & Nieuwenhuis, S. (2014). Moral concerns increase attention and response monitoring during IAT performance: ERP evidence. *Social Cognitive and Affective Neuroscience, 9,* 141-149. https://doi.org/10.1093/scan/nss118

Vergoossen, H. P., Renström, E. A., Lindqvist, A., & Gustafsson Sendén, M. (2020). Four dimensions of criticism against gender-fair language. *Sex Roles, 83*, 328-337. https://doi.org/10.1007/s11199-019-01108-x

Vincent, B., & Manzano, A. (2017). History and cultural diversity. In *Genderqueer and non-binary genders* (pp. 11-30). Palgrave Macmillan, London. https://doi.org/10.1057/978-1-137-51053-2_2

Vink, M., Derks, B., Ellemers, N., & van der Lippe, T. (2023). Penalized for challenging traditional gender roles: why heterosexual relationships in which women wear the pants may be more precarious. *Sex Roles, 88*(3-4), 130-154. https://doi.org/10.1007/s11199-022-01339-5

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of Personality and Social Psychology, 54*(6), 1063-1070. https://doi.org/10.1037//0022-3514.54.6.1063

Webster, S. W., & Albertson, B. (2022). Emotion and politics: Noncognitive psychological biases in public opinion. *Annual Review of Political Science, 25*(1), 401-418. https://doi.org/10.1146/annurev-polisci-051120-105353

Weißflog, M. I., & Grigoryan, L. (2024). Gender categorization and stereotypes beyond the binary. *Sex roles, 90*(1), 19-41. https://doi.org/10.1007/s11199-023-01437-y

Wells, J. C. (2007). Sexual dimorphism of body composition. *Best Practice & Research Clinical Endocrinology & Metabolism, 21*(3), 415-430. https://doi.org/10.1016/j.beem.2007.04.007

Wickham, M. I., van Nunspeet, F., & Ellemers, N. (2023). Gender Identification Beyond the Binary and Its Consequences for Social Well-Being. *Archives of Sexual Behavior, 52*(3), 1073-1093. https://doi.org/10.1007/s10508-022-02453-x

Wong, C. Y. E., Kirby, T. A., Rink, F., & Ryan, M. K. (2022). Intersectional invisibility in women's diversity interventions. *Frontiers in Psychology, 13*. https://doi.org/10.3389/fpsyg.2022.791572

World Health Organization (2022). https://www.who.int/news/item/06-07-2022-who-updates-widely-used-gender-mainstreaming-manual

Worthen, M. G. (2022). This is my TERF! Lesbian feminists and the stigmatization of trans women. *Sexuality & Culture, 26*(5), 1782-1803. https://doi.org/10.1007/s12119-022-09970-w

Yang, M. (2021, October 7). 'I'm team TERF': Dave Chapelle under fire over pro-JK Rowling trans stance. The Guardian. https://www.theguardian.com/stage/2021/oct/07/dave-chappelle-transgender-netflix-special-backlash

Yeung, S. P., & Wong, W. I. (2018). Gender labels on gender-neutral colors: Do they affect children's color preferences and play performance?. *Sex Roles, 79*(5), 260-272. https://doi.org/10.1007/s11199-017-0875-3

Zlotnick, S. (2001). *Women, writing, and the industrial revolution*. JHU Press. https://doi.org/10.56021/9780801858291

Acknowledgements

Writing this thesis was a monumental task, the largest task I have completed in my life thus far. It was a journey with many ups and downs, and I could not have done it without the support of wonderful people. I would like to dedicate this space to highlighting a few of those people.

First, I would like to thank my supervisors, **Félice van Nunspeet** and **Naomi Ellemers**. Naomi kept me sharp and challenged me to think beyond the confines of my own ideas and beliefs; through her I was always reminded of the big picture of what I was trying to achieve. Félice, on the other hand, helped me see how sometimes details matter and how readers may interpret them. More importantly though, as my daily supervisor she was a source of emotional support and safety whenever I felt like the "birth" of this thesis was going to drive me up the wall. Between the two of them, they gave me the tools I needed to become the researcher I am today.

Next, I would like to thank my paranymphs, **Kshitij Mor** and **Onur Şahin**. Kshitij, we knew we needed to be friends the moment we first met in that fateful Teams meeting (ah, covid times). I am extremely grateful that you joined the research group and that I got to meet you. My designated party buddy: I hope we spend many more nights at Bodytalk together, or just at one another's houses, gossiping over gin and tonics. I am sure you will help me make the party for this defence a night to remember, too! Onur, we have been 'roommates' at the Langeveld building since the beginning of both of our PhDs. Together, we have celebrated academic victories, lamented the sometimes less joyful sides of academia and fought for more inclusion at our own Faculty. I am incredibly thankful that I get to walk you down the academic aisle, and you me.

I would also like to thank the Assessment Committee **Jojanneke van der Toorn, Colette van Laar, John de Wit, Florien Cramwinckel**, and **Daan Scheepers**. Thank you for your time, your careful consideration, your comments, and critical questions.

Thank you to the person who designed the cover of my thesis, and helped me format it: **Aïsha Kanters**.

Many thanks to my (former) academic colleagues, whom I could always count on for support in the form of cups of coffee, sarcasm, rants and venting, Jenga games, a shoulder to cry on, and funny stories about tomatoes, monkeys, or the colour indigo. For instance, I would like to thank **Onur Şahin**, **Kshitij Mor**, **Nadia Buiter**, **Gonneke Ton**, **Esmee Veenstra**, **Piet Groot**, **Marty Colombo**, **Johanna Kruger**, **Frank Gootjes**, **Ilona Domen**, **Larisa Riedijk**, **Melissa Vink**, **Elena Bacchini**, **Eva Grosfeld**, **Jeanette van der Lee**, **Savanne Grotenhuis**, and **Martine Kloet**, to name a few.

Of course, I would like to thank the entire **Organisational Behaviour lab**, current and former members. Many warm thanks to **Karin Dirks-Hansen** and **Annemarieke Blankesteijn** for answering all of my annoying administrative questions. A special thanks to **Jojanneke van der Toorn**, who made time to have a casual coffee with me on a regular basis, to brainstorm about gender and LGBTQI+ topics. You were a sort of advisor and mentor to me at times, and I deeply appreciated it.

Besides my direct (former) colleagues, I would also like to thank everyone I met at conferences and courses within, for instance, the **KLI**, **ASPO**, **EASP**, and the **Gender & Diversity Hub** of Utrecht University. Countless people have inspired me with their talks, posters, and through networking during lunches and coffee breaks. A special thanks to colleagues at KU Leuven, particularly **Colette van Laar** and **Jenny Veldman**, who welcomed me for an (online) academic visit. I wish it could have been in person, but the "global event that shall not be named" did not permit it. I also would like to give a special shoutout to my colleagues at **Tijdschrift voor Genderstudies** – working with you on the journal has been a joy!

Veel dank ook aan mijn nieuwe collega's van **ZINZIZ**, met hun luisterend oor wanneer ik even stoom moest afblazen over het proces rondom het schrijven van mijn proefschrift. Jullie hebben elk compleet hoofdstuk met me gevierd tijdens de lunch overleggen op vrijdagen,

en met me gebrainstormed over hoe ik de kennis vanuit mijn PhD kan gebruiken voor het soort onderzoek dat ik nu met jullie doe.

To my friends who were outside of my direct academic bubble, I want to thank you all. Special mentions go to **Nadia Buiter** (whom I cannot only include in the list of former colleagues as that would not do you enough justice), **Martijn Clausen** and **Iris ten Hoopen** (thank you for the Korean food and board games).

Ein riesiges Dankeschön an meine Familie für die bedingungslose Unterstützung und Ermutigung auf meiner akademischen Reise. Ihr habt mich zum ehrgeizigen und sturen Mensch gemacht, der ich heute bin. "Wenn du ein Tankwärter sein willst, sei der beste Tankwärter der du sein kannst", haben **meine Eltern** immer gesagt. Ich bin kein Tankwärter geworden, aber ich arbeite jeden Tag daran, die Beste Forscherin zu sein, die ich sein kann. Meine liebe **Oma**, wessen Namen ich als Mittelnamen habe. Du bist nicht mehr bei uns in der physischen Welt, aber du bist immer bei mir im Geist. Meine **Familie in Hamburg und Umgebung**: ich danke euch für eure Gastfreundschaft und eure Geduld in den letzten Jahren, in denen ich sehr viel weniger Zeit hatte um euch zu besuchen.

To my British family, who welcomed me into your **Griffel-Wick-Foot clan** as soon as I entered Rob's life, I am overwhelmed by how you have "adopted" me and will always be grateful for that. Jo, our chats about being a working woman have inspired me throughout this PhD.

Last, but certainly not least, to **Rob**. I could never thank you enough. I would especially like to thank you for cups of tea and food when I have been hyper-focusing and don't even know who I am anymore (also known in our household as a state of "emg"). You have been a lifesaver throughout this process.

I could name countless more names, and if I have forgotten you and you feel you should be on this list, you have my permission to send me a strongly worded letter. In fact, I would welcome it!

Curriculum Vitae

Miriam Wickham was born 18th September 1993 in Oberhausen, Germany. Her parents quickly moved her to Los Realejos on the island of Tenerife, Spain. She spent her first few years in Spanish state schools, growing up bilingual. At the age of 11, she joined a British school on Tenerife, which paved the way for her to complete her Bachelor's degree In Psychology and Philosophy at the University of Reading, UK, in 2014. She moved to the Netherlands for her Master's in Neuroscience and Cognition at Utrecht University, which she completed in 2016. She spent one year working as a research assistant for the Organisational Behaviour group of Utrecht University, under Prof. Dr. Naomi Ellemers, before beginning her PhD at this group in 2017.

Miriam is currently working at ZINZIZ bv in Utrecht as a researcher. At this company, she conducts qualitative and quantitative research into issues surrounding (labour) participation, often paid for by government agencies such as UWV, ministries or local municipalities. In her position, she works as executive researcher, and is also project lead on some projects.

Journal Publications

Wickham, M. (2020). Open science and the exaggeration of gender differences in scientific literature. *Tijdschrift voor Genderstudies, 23*(2), 139-142.

Sternheim, L. C., **Wickham, M. I.**, Danner, U. N., Maddox, T. W., Filoteo, V. J., Shott, M. E., & Frank, G. K. (2021). Understanding implicit and explicit learning in adolescents with and without anorexia nervosa. *Journal of Eating Disorders, 9*, 1-15.

Wickham, M. I., van Nunspeet, F., & Ellemers, N. (2023). Gender identification beyond the binary and its consequences for social well-being. Archives of Sexual Behavior, 52(3), 1073-1093.

van Eersel, J. H., Hulshof, I. L., **Wickham, M. I.**, Smid, G. E., & Boelen, P. A. (2024). Pre-job loss grief reactions and work attachment among sick-listed employees: Introduction of the imminent Job Loss Scale. *BMC Psychology, 12*(1), 118.

KLI Dissertation Series

The "Kurt Lewin Institute Dissertation Series" started in 1997. The following dissertations have been published during the last two years. The complete list can be found on our website:
https://kurtlewininstituut.nl

2022-01: **Iris van Sintemaartensdijk**: Burglary in virtual reality

2022-02: **Lu Liu**: Public participation in decision making on sustainable energy transitions

2022-03: **Rabia Kodapanakkal**: The role of tradeoffs and moralization in the adoption of big data technologies

2022-04: **Elissa El Khawli**: Why, When, and How Workers Regulate: A lifespan perspective on work design and emotion regulation at work

2022-05: **Chantal van Andel**: Clinical Grade Differences Between Ethnic Minority and Majority Students: Institutional-, assessor and student-related factors

2022-06: **Inga Rösler**: Hear me out: How to create an open mind towards moral criticism

2022-07: **Tessa Coffeng**: Bias in supervision: A social psychological perspective on regulatory decision-making

2022-08: **Babet Kanis**: Hope and Health in the Face of Adversity

2022-09: **Martijn Blikmans**: Do we live in the age of emotion politics? The effects of anger, disgust, hope, and nostalgia communication on political support and polarization

2022-10: **Anne van Valkengoed**: Reality, causes, consequences: The role of climate change perceptions in climate adaptation

2022-11: **Dan Sun**: How People Learn to Act on Goals: A New Examination of the Mechanistic Ideomotor Action Account

2022-12: **Carla Roos**: Everyday Diplomacy: Dealing with controversy online and face-to-face

2022-13: **Christhian Martínez**: Hate: Distinctive Features Across Individuals and Groups

2022-14: **Wenrui Cao**: Forgiveness in Work Relationships: Causes and Consequences

2022-15: **Jiafang Chen**: How Narcissists Navigate the Communal World

2023-01: **How Hwee Ong**: Demystifying Magical Justice Beliefs: Believing in Justice in a World of Injustices

2023-02: **Erdem Meral**: Talking about belonging: Whether, why, and how people talk about social exclusion

2023-03: **Maria Zwicker**: The Complexity of Consumer Attitudes Towards Sustainable Alternatives

2023-04: **Shuxian Jin**: Social Dilemmas and Institutional Solutions to Promote Cooperation

2023-05: **Frank Gootjes**: Societal discontent as a catalyst for action: Explaining protest and solidarity, why we help and protest

2023-06: **Nieke Lemmen**: The autonomy challenge, 5 ways to control control

2023-07: **Leonie Vrieling**: When others are in control: Understanding public responses to externally controlled energy projects

2023-08: **Piet Groot**: Born abroad, studied here: A historical and psychological account of migrant doctors integrating

2023-09: **Kunalan Manokara**: Expressing positive emotions as they are: Spontaneous Production, Display Rules, and the Role of Culture

2023-10: **Lei Fan**: Navigating Threats Through Disgust: From Pathogen Avoidance to Moral Punishment

2023-11: **Kaiyang Qin**: Revisiting Automatic Goal Pursuit: Exploring the Value of Goals in Cue-Based Behavior

2023-12: **Xiao Wang**: A tale of two identities: The role of environmental self-identity and environmental group-identity in motivating pro-environmental behaviour

2023-13: **Yingqiu Wu**: Cancer in the Media: Unintended Effects on

Prevention

2023-14: **Valentina Lozano Nasi**: Human Transilience in the Face of Adversities: Embracing Global Challenges as Stepping Stones, rather than Stumbling Blocks

2023-15: **Ana Leal**: An Intergroup Perspective on Attitude Moralization

2023-16: **Monique Chambon**: The Complexity of Behaviour in relation to Health, Safety and Sustainability: A Psychological Network Approach

2023-17: **Terence Daniel Dores Cruz**: The Cooperative and Competitive Functions of Gossip

2024-01: **Çağla Çınar**: Adapting Food Preferences: The Social, Emotional, and Genetic Roots of Preferences for Meats versus Plants

2024-02: **Judith Langerak**: One step ahead: Proactive coping to minimize job insecurity

2024-03: **Leon Hilbert**: Decisions under financial scarcity

2024-04: **Gonneke Ton**: Between the Devil and the Deep Blue Sea? Exploring the unique experience of ambivalence in the context of societal debates

2024-05: **David Schultner**: The Transmission of Social Biases through Instrumental Learning

2024-06: **Annika Karinen**: The social, political, and aesthetic implications of disgust

2024-07: **Miriam Ines Wickham**: Square Peg in a Round Hole: Gender Beyond the Binary

Colofon

ISBN: 978-90-393-7718-5

DOI: https://doi.org/10.33540/2429

Cover design & layout : Aïsha Kanters | www.aishakanters.com

Printing: Ipskamp Printing | www.ipskampprinting.nl