Powerful males on top: Spatial simulations in the mental representation of gender stereotypes

Natalia Zarzeczna

A thesis submitted in partial fulfilment of the requirements

for the degree of Doctor of Philosophy

Supervisors: Ulrich von Hecker and Geoffrey Haddock

School of Psychology, Cardiff University

October 2018

CARDIFF UNIVERSITY

PRIFYSGOL CAERDYDD

School of Psychology

DECLARATION

This work has not been submitted in substance for any other degree or award at this or any other university or place of learning, nor is being submitted concurrently in candidature for any degree or other award.

Signed ………………………………………… (candidate) Date …………………………

STATEMENT 1

This thesis is being submitted in partial fulfillment of the requirements for the degree of …………………………(insert MCh, MD, MPhil, PhD etc, as appropriate)

Signed ………………………………………… (candidate) Date …………………………

STATEMENT 2

This thesis is the result of my own independent work/investigation, except where otherwise stated. Other sources are acknowledged by explicit references. The views expressed are my own.

Signed ………………………………………… (candidate) Date …………………………

STATEMENT 3

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed ………………………………………… (candidate) Date …………………………

STATEMENT 4: PREVIOUSLY APPROVED BAR ON ACCESS

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loans **after expiry of a bar on access previously approved by the Academic Standards & Quality Committee.**

Signed ………………………………………… (candidate) Date …………………………

Summary

The theory of grounded cognition proposes that abstract thinking involves mental simulations of acquired sensorimotor experiences. In this thesis, my primary aim was to investigate whether the mental representation of gender stereotypes is related to sensorimotor experiences associated with space. In a series of experiments, I asked participants to complete a spatial task and categorise as quickly as possible which person (one female name; one male name) of two vertically displayed on the screen (top and bottom) was socially powerful or powerless. As predicted, stereotypically constructed power-gender associations involved simulations of males as powerful at the top as opposed to females as powerful at the top. This was specific to spatial location, as males as powerful, but not females, were chosen faster at the top than at the bottom. Further, I replicated the above findings using pupillometry. Additional mechanisms as well as consequences of spatial simulations were further examined. I found that spatial simulations in gender-power associations are dependent on simultaneous power and gender salience. Finally, I demonstrate that spatial features involved in the representation of gender stereotypes have consequences for gender perceptions by promoting stereotype-consistent thinking.

Altogether, my studies show that the representation of power and gender involves spatial simulations that are influenced by stereotypic thinking. When making judgements about power differences, the spatial cues and socially constructed stereotypical beliefs compete for activation to accomplish conceptual integration and produce meaning. Such findings imply that mental representations of stereotype-consistent knowledge are more pronounced as they involve concrete spatial simulations.

Acknowledgments

First of all, I would like to thank my supervisors, Ulrich von Hecker and Geoff Haddock, for their interest in my project and agreeing to supervise me. Their expertise and guidance in combination with their eagerness to listen and acknowledge my ideas significantly helped me develop as a confident researcher. I am very grateful for their support and clear communication in delivering feedback and discussing literature. Thanks to this, I always found our meetings motivating and academically enriching. I am certain that the knowledge and skills they helped me acquire will be essential in my future career.

Second, I wish to thank my friends, and colleagues from the Social Psychology lab who provided me with a very supportive environment during my studies. Specifically, I would like to thank: Christina[1] and Lukas (Sosayetti), Paul, Nas, and Geoffrey for their eagerness to talk about my project (even during our holidays) and always being enthusiastic to hear about the results from my experiments; Kelsey for being very supportive, especially in the last year; Travis for introducing me to pupillometry and coding; Kasia, Karis, and Suzanna for always being encouraging.

Również chciałabym podziękować rodzicom za wsparcie w moim wyborze kariery.

Finally, I wish to thank ESRC for funding me, and all the research assistants (especially Anna Błaszczak and Paul Goesmann) who helped in data collection in the UK and Poland.

[1] Importantly, I would like to thank Christina for the 68 and her feedback on how to talk about spatial cognition to "recyclers."

Table of contents

Chapter 1: General Introduction[2]

Overview

In this chapter, I review the literature on grounded cognition as well as discuss research and theories concerning spatial simulations. Subsequently, I present currently available knowledge about mental representations of gender stereotypes and summarise research on social power. Finally, I integrate the findings on grounded cognition and spatial simulations with research on stereotypes and power to present the rationale behind investigating spatial simulations in gender stereotyping. At the end of the chapter, I provide an overview of my studies and discuss potential implications of my project.

"Space: Literally it means *nothing*, a vacuum between stars and planets, but by the same token it means *everything*. It's what connects all our worlds." - Captain Janeway, Star Trek Voyager[3]

While physical space exists in the distance between objects, space is also a medium that connects abstract ideas within our *mental* worlds. It provides the context for our daily lives and interactions with other individuals. From our early days, humans are exposed to physical spatial distances and orientations, and experience them through perceptual and motor capabilities. Such experiences shape human physical and mental development (Smith & Gasser, 2006). Indeed, numerous studies and social cognitive theories suggest that abstract cognitive processing is grounded in sensorimotor experiences (Niedenthal, Barsalou, Winkielman, Krauth-Gruber, & Ric, 2005;

---

[2] This introductory chapter is partly based on two manuscripts in preparation (Zarzeczna, von Hecker, & Haddock, 2018a, 2018b)

[3] The quote was obtained from: Quotes. (n.d.). Retrieved June 18, 2018, from https://www.imdb.com/title/tt0708942/quotes

Williams, Huang, & Bargh, 2009), and that in this way, abstract concepts (e.g., power) are likely to be understood by means of concrete concepts (see Bargh, 2006; Barsalou, 2003; Clark, 1973; Lakoff & Johnson, 1999; Mandler, 1992). Lakoff and Johnson (1999) further argue that abstract concepts can be illustrated by concrete concepts in linguistic metaphors. Literally, bigger and stronger animals are more likely to win a fight, and those who win are on top of their opponents. The winners, in turn, become powerful within the social hierarchy. Hence, metaphorically, *physical size* is linked with *being on top,* implying greater power. In terms of social relations, power is commonly conceptualised as an ability to control or influence others (see Guinote, 2017, for an overview). Lakoff & Johnson (1999) suggests that literally it is easier to control/exert pressure or influence someone when we are physically positioned at the top as opposed to the bottom. Consistent with this framework, Jamrozik, McQuire, Cardillo, and Chatterjee (2016) claim that metaphors do not simply reflect thought processes about linking abstract concepts with concrete ones, but serve as a tool that guides the development of the links between embodiment and abstraction. To summarise, space might not only connect all our *physical* worlds, but it also seems to connect abstract ideas in our *mental* worlds.

Spatial simulations of power

In the second half of the twentieth century, cognitive scientists did not consider the potential significance of the interrelation between human mind and sensorimotor experiences. Due to advances in computer science, mental processes, such as reasoning or thinking about abstract concepts, were seen as an array of codes or symbols functioning like computer software (see Shapiro, 2011, for an overview). However, by integrating empirical evidence from cognitive and social cognitive experiments, Barsalou (1999) suggested that the mental representation of abstract concepts involves re-enactment of previously acquired perceptual/motor experiences (i.e., perceptual symbols). This process is also referred to as *simulation.* The perceptual symbols, grounded in experiences, were no longer seen as being completely abstract. For example, regarding the concept of power, this theory suggests that thinking about power likely involves a simulation of associated and memorised perceptual symbols of space. As implied by this example, simulation involves a mental representation of a concept that makes use of knowledge based on sensorimotor experiences with a physical dimension. The sensorimotor experiences that conflate power with verticality might be based on human relations formed early in life (see Schubert, 2005, for an overview). Specifically, through interactions with their parents, children notice that adults are taller and have more power than them. Therefore, children learn that having more power involves being taller. In this way, a perceptual symbol, whereby power is associated with upper vertical positions (i.e., tall-powerful parents), is created (Schwartz, Tesser, & Powell, 1982). The perceptual symbol of space and power is reinforced further during adulthood. Research indicates that physical height is correlated with social status and power. That is, taller individuals tend to have a higher social-esteem (i.e., more positive social evaluations), they are more likely to emerge as leaders, and they enjoy greater success at work, as compared to physically shorter individuals (Judge & Cable, 2004). Finally, Fiske (2004) demonstrated that people use spatial language (high, up, top) in order to determine power differences among individuals, whereby physical distances between vertical positions indicate power differences (see Giessner & Schubert, 2007). Specifically, placing an individual in a higher position on the vertical dimension indicates that the individual has more power than those placed below them. Overall, it appears that early social interactions with powerful and physically tall adults form the basis of associations between space and verticality. Such associations are further reinforced by language and our perception of the social world where tall people are more successful than short people. To illustrate how such perceptual symbols function in cognition, below I provide an example of the involvement of verticality cues in the mental representation of power.

The idea that spatial simulations take place in the representation of power is supported by empirical studies. Schubert (2005) demonstrated that power primes higher vertical positions, whilst lack of power primes lower vertical positions. In his research, participants indicated as fast as possible which group, as displayed vertically on a computer screen (top versus bottom), was the powerful or powerless in a pair (e.g., *master* versus *servant*). The vertical position of the groups was counterbalanced across trials. Participants were faster at responding to powerful groups that appeared at the top (compared to bottom) while looking for powerful groups, and faster at responding to powerless groups when they appeared at the bottom (compared to top) while looking for powerless ones. This pattern indicated a congruency between participants' mental representation of power-related concepts and the physical spatial locations of these concepts. That is, when mentally representing abstract concepts (power), people simulate the concepts in the associated spatial locations derived from sensorimotor experiences (see Giessner & Schubert, 2007). Similar findings have been obtained in the context of dominance (von Hecker, Klauer, Wolf, & Fazilat-Pour, 2016), social status (von Hecker, Klauer, & Sankaran, 2013), valence (Meier & Robinson, 2004), and agency (Suitner & Maass, 2016). Such embodied simulations take place automatically when people represent abstract concepts (i.e., power or valence) associated with sensorimotor experiences (verticality). Therefore, it is different to other forms of simulations, namely, mental imagery or mechanical reasoning. This is because these forms are deliberate processes of representing how objects move in working-memory. Such simulations are used to solve a given problem (e.g., mental rotation, see Hegarty, 2004 for simulations in mechanical reasoning, and MacInnis & Price, 1987, for mental imagery). Further, it is important to note that mental representation is an internal mental image of an acquired concept that may or may not use spatial simulations. Specifically, spatial simulation used in a mental representation of abstract concepts would involve a mental positioning of those concepts on a spatial dimension.

In addition, evidence from cueing paradigms also indicates that the processing of concepts associated with power directs attention to vertical positions. Zanolie et al. (2012) asked participants to complete an attention-cueing task. The task was to identify as quickly as possible the location of a target stimulus on a computer screen. However, before the presentation of the target stimulus, participants were primed with power-related concepts (one concept at a time; e.g., king or servant) and asked to determine whether the concept represented a powerful or powerless person. They found that participants were faster to identify the target when it was presented in the upper vertical position after being primed with powerful individuals as opposed to powerless ones. The opposite was true for the powerless groups and bottom vertical positions. Similar findings were obtained for other concepts. For example, in a study by Taylor, Lam, Chasteen, and Pratt (2015), when participants were primed with positive self-esteem traits, they were faster to detect a target stimulus at the top as opposed to the bottom. The same was observed for negative traits presented at the bottom. Interestingly, these results only occurred when the targets were presented on the vertical dimension, but not the horizontal dimension. Also, semantics was essential for the effect to occur, as superficial processing of the cues (i.e., considering orthography versus valence) did not produce verticality effects. Altogether, research indicates that reasoning about abstract concepts, such as power, involves activation of sensorimotor experiences associated with the vertical spatial dimension.

Spatial simulations in conceptual integration

The spatial simulation process described above is relevant to work on conceptual integration or blending, proposed by Fauconnier and Turner (1998). Specifically, conceptual blending refers to a process whereby semantics of two or more *consistent* concepts are combined or conflated. This process results in a combined meaning of the blended concepts. To illustrate this process, Casasanto (2009, pp. 362) provided an example of blending among primacy, left horizontal position, and goodness, in which semantics of these concepts are combined: "Linguistic expressions like "the prime example" conflate primacy with goodness (i.e., this phrase can mean the first example, the best example, or both). Speakers of languages like English may be predisposed to consider the leftmost item to be the first and therefore the best. This metaphorical blend of left, first, and best should result in a culturally constructed Good Is Left bias." The concepts that can be combined in a mental representation to produce meaningful outcome are defined as "blendable," that is, their semantics overlap. Further, relevant to the work on grounded cognition, Fauconnier and Turner (1998) argue that simulations are not merely mental representations of integrated or blended abstract concepts, but in fact, the simulations can actively assist in the blending or integration of concepts and serve meaning production. Such blending and involvement of spatial elements is again dependent on the compatibility or consistency between the concepts. For instance, conflating concepts whose mental representations involve spatial features (primacy) with non-spatial concepts (goodness) produces a blend where goodness is represented on the left. Likewise, von Hecker et al. (2016) demonstrated that dominance (in the sense of abstract magnitudes, e.g., "older than," "richer than") is blended with primacy and hence associated with the left side in cultures with left-to-right writing/reading habits, whilst with the right side in cultures with right-to-left habits, where primacy is associated with the right. To summarise, it is likely that blending of concepts might involve spatial processing as long as one of the blended concepts is grounded in perceptual experiences with space.

Social beliefs in conceptual integration

Extant literature, as discussed above, has focused on spatial simulation of generic concepts that have strong power-related implications. That is, in Schubert's (2005) research, a long-term memory connection between presented stimuli and power via the literal meaning of the stimulus (e.g., boss, subordinate) was likely responsible for spatial simulations of powerful-top and powerless-bottom, as the *boss* has the ability to control the *subordinate.* Yet, in real contexts of thought and conversation, concepts vary in the degree to which they are associated with power. Many concepts are malleable with respect to their associations and implications. For example, "master" in relation to "servant" likely carries stable power implications and therefore may be reliably simulated in terms of vertical space ("master" = top, "servant" = "bottom"). On the other hand, whether or not the relation between specific individuals or group members is simulated in vertical mental space may be more malleable and dependent on whether both concepts are seen in a particular power relation or not. If they do, I argue that this will be a case in which a blending takes place between concepts that are reliably simulated in vertical space ("high power" vs. "low power") and concepts that are simulated in vertical space only if the context is semantically relevant (e.g., when seen in a power context; see Taylor et al. 2015). Likewise, Fauconnier and Turner (1998) further suggest that blending between concepts, that is, their compatibility, may also depend on personal beliefs. As the main aim of the present thesis, I wish to demonstrate that personal social beliefs, such as stereotypic associations, can have the effect of eliciting a blend, and therefore initiating spatial simulation of concepts that are otherwise not necessarily simulated spatially.

Mental representation of stereotypes

In the current thesis, I focus on investigating the precise features in the mental representation of gender stereotypes in relation to social power. First, in this section, I stress the importance of studying gender stereotypes associated with power in the context of their undesirable social consequences, such as prejudice. Here, I present initial research indicating that stereotypes have stable characteristics (they are activated across contexts). Second, I present more recent research findings showing that stereotypes are situated. In this context, I suggest that social power can be considered as one of the attributes of gender that provides salient and diagnostic information when categorising and differentiating males from females in relevant situations (Ellemers, 2018; Ford & Stangor, 1992).

Despite the fact that many research studies have been conducted to broaden our understanding of stereotypes, it seems that stereotypes still represent a social problem. Biased beliefs about social groups still shape people's judgements of others. For example, undergraduates in biology tended to overestimate the academic performance of male students, despite the fact that female students had higher grades. This bias was only demonstrated by males (Grunspan et al., 2016; see also Leslie, Cimpian, Meyer, & 2015; Moss-Racusin, Phelan, & Rudman, 2012). Also, female university professors are awarded fewer prestigious positions than males, even though they exhibit equal levels of academic performance relative to males (Trevino, Gomez-Mejia, Balkin, & Mixon, 2015). Another recent study indicated that people evaluate teachers' performance more positively when the teacher is identified by a male rather than a female name (MacNell et al. 2015).

Furthermore, gender stereotypes seem to be linked to prejudice such as sexism. While stereotypes are incorrect beliefs about social groups functioning on the level of cognitive associations, prejudice is a biased and typically negative attitude towards a group member based on faulty premises (Allport, 1954). That is, prejudice represents an overall negative attitude towards an individual belonging to a stigmatised group. Although initial research indicated that stereotypes are only modestly related to prejudice (r = .25; Dovidio, Kawakami, Johnson, Johnson, & Howard, 1997), studies on sexism suggest that sexist attitudes are associated with stereotypes that can be both negative, but also ostensibly positive in nature (Katz & Hass, 1988; Glick & Fiske, 1996). That is, the endorsement of ostensibly positive beliefs about women having communal traits complements the endorsement of negative beliefs suggesting that women do not possess the intellectual skills needed to succeed in a professional career. Both types of stereotypes are linked with overall prejudice against women, characterising them as weak and incompetent. Such prejudice, in turn, might lead to discrimination against women (e.g., lower employability; see Ellemers, 2018; Glick & Fiske, 1996; Kawakami, Dion, & Dovidio, 1998). To summarise, stereotypes seem to be associated with biased and inaccurate stable judgements about females. While such inaccurate beliefs impact social judgements, they are also linked with an overall negative evaluation of females in the form of prejudice that has been further associated with discrimination. This indicates that it is important to identify ways to reduce stereotyping. To achieve that we need to better understand when stereotypes are likely to be activated and applied. It is important to note that not all stereotypic attributes are salient, activated, and applied to judgements of stereotyped groups at the same time and across all situations. Below, I explain why the attribute of social power is relevant in gender stereotyping and when it is the most likely to be applied to judgements about gender.

First, using a stereotypic attribute (social power) to distinguish social groups (gender) represents an efficient way of processing and understanding the complexity of the social world (Macrae & Bodenhausen, 2001). Even though stereotypic judgements are inaccurate (Ellemers, 2018), Fiske and Taylor (2010) suggest that when evaluating others, people are not motivated to engage in deep mental processing. Instead, people prefer to rely on shortcuts, e.g., stereotypes, that are less cognitively demanding. This functionalist approach to stereotyping suggests that using stereotypic attributes in social categorisations is mentally beneficial, providing quick and efficient judgements. Therefore, assessing an individual's level of social power on the basis of their gender would involve engaging fewer cognitive resources than analysing a number of attributes represented by that individual.

Second, people might be motivated to categorise males and females according to the stereotype of social power due to its salience across social contexts. This is because within the Western societies, men occupy more powerful social roles than women (Ellemers, 2018). Further, people endorse ostensibly positive beliefs that women have communal traits, e.g., they are caring or kind, so they are good mothers and wives. In turn, such beliefs might be linked with negative stereotypes suggesting that women would underperform in more powerful roles, such as leadership or career development (Glick & Fiske, 2001). Both forms are complementary in maintaining social inequality at workplace, with men still occupying the most powerful social positions (Carli & Eagly, 2001; Ellemers, 2018; Rudman & Kilianski, 2000; Rudman & Phelan, 2010). Such beliefs are present, even though it appears that cognitive performance or leadership skills do not differ between groups of men and women (Hyde, 2014; see Ellemers, 2018, for an overview).

Finally, while people might be prone to automatically activate the stereotype of power and gender and apply it to their judgements, the concept of power might not always be activated in the mental representation of gender. As mentioned earlier, although it is likely that the concept of power is used in the representation of gender stereotypes, this will only be the case when the attribute of power provides relevant and essential information that would differentiate males from females. This notion stems from the research on central tendency and variability in stereotypic attributes. For example, Ford and Stangor (1992) showed that people use the central tendency and variability of stereotype-relevant group attributes in order to evaluate that group. In their experiment, participants were presented with behaviours of individuals belonging to two groups (i.e., blue and red). The two groups differed on two dimensions in terms of friendliness and intelligence ratings. However, the extent of such differences was manipulated in terms of the mean ratings. The difference in mean rating between the blue and red group on one dimension (friendliness in one condition; intelligence in the other) was always higher than the difference in mean rating between the two groups on the other dimension (intelligence and friendliness, respectively). That is, one dimension represented larger differences in mean ratings than the other dimension. When participants were then asked to characterise the two groups, in both conditions they relied more on the dimension that represented the larger mean differences in comparison to the smaller mean differences. Further, Ford and Stangor (1992) found that when characterising social groups, their participants tended to use attributes that were associated with smaller, rather than larger, within-group variability. Overall, this research indicates that when forming impressions of social groups, people tend to rely on attributes that (a) are associated with the largest differences between the groups and (b) denote small within-group variability. Sani and Bennett (2001) provided further support for such findings by showing that 7-year-old children's use of gender stereotypes when evaluating their in-group depended on the comparison context. For example, boys were more likely to categorise their in-group as brave and strong only after being asked to evaluate *girls*, but not *men* (see also Bennett & Sani, 2006). In line with the above findings, Lebois, Wilson-Mendenhall, and Barsalou (2015) demonstrated that spatial attributes (i.e., top versus bottom) associated with words such as *sky* or *basement* are activated only in relevant contexts, when the spatial location is salient. Such findings further suggest that it is unlikely that concepts carry stable mental representations that are activated in all contexts.

Research on multiple categorisations also supports the notion that stereotypic features of a social group are activated only in relevant conditions. Hall and Crisp (2005) asked their participants to consider groups that students may belong to other than their university membership. They found that when students considered such alternative affiliations, they were less likely to exhibit in-group favouritism when evaluating students from their own university (in-group) compared to the students from another university (out-group). To summarise, this research suggests that the attributes activated in mental representations of gender stereotypes will depend on the context in which a judgement about males and females is being made.

The literature discussed above indicates that stereotypic attributes associated with social groups are not fixed or stable mental structures. This implies that stereotypes might not be automatically activated or applied to judgements under all circumstances. On the one hand, initial research on stereotype activation proposed that stereotypes are automatically activated upon encountering a member of a stereotyped group. Devine (1989) demonstrated that stereotypes are characterised by two components, automatic and controlled. Because most people have knowledge about stereotypes, she suggested that stereotypes are automatically activated when a member of a stereotyped group is encountered. Only people who are not prejudiced are able to inhibit such automatically activated stereotypes and report unbiased thoughts. Subsequent research indicated that in fact only highly prejudiced individuals automatically activate stereotypic content upon encountering a group member, but that all individuals (regardless of level of prejudice) evaluate stereotyped individuals negatively after direct priming with a negative stereotype (Lepore & Brown, 1997). On the other hand, recent models of stereotyping emphasise the importance of several factors that contribute to the stereotype activation, application, and inhibition. These factors are associated with person's goals, the level of prejudice, available cognitive resources, and learned associations (Gilbert & Hixon, 1991; Kawakami, Dovidio, Moll, Hermsen, & Russin, 2000; Lepore & Brown, 1997; Sinclair & Kunda, 1999). Kunda and Spencer (2003) further proposed that the stereotype activation and application depend on comprehension goals (i.e., the need to simplify events and understand them), self-enhancement goals (i.e., the need to enhance or affirm self-esteem), and the motivation to avoid prejudice. Each of these factors might inhibit or encourage stereotyping in situations where stereotype activation and application enable a goal achievement. For example, if stereotyping promotes the goal of enhancing self-esteem via in-group favouritism, aids the understanding of complex events, and there is a low motivation to avoid prejudice, then people would be more likely to activate and apply (rather than inhibit) a stereotype. Which of these goals is salient depends on situational factors that might promote a specific goal activation. Such activation might be either spontaneous and automatic or deliberately activated due to certain thoughts. To summarise, the model clearly indicates that the stereotype activation, application, or inhibition are determined by situational factors.

In summary, the presented research suggests that the content and use of stereotypes in social judgement is based on personal goals and the functionality of stereotypes in a given situation. Therefore, stereotypes do not carry stable mental representations that are activated and applied automatically across different contexts. Instead, mental representations of stereotypes are malleable. Therefore, in my own research, I investigated the features associated with the mental representation of gender stereotypes in contexts where the attribute of power is salient and can serve the function of distinguishing males from females.

The above insights on stereotyping broadens our knowledge of how stereotypes impact people's behaviour and why it is important to reduce their impact. However, there is also a need to investigate the mechanisms and representational features that might be associated with specific stereotypes. Therefore, one of the aims in my thesis is to address this issue. People develop stereotypes over time due to socialisation but once stereotypes are acquired, they are usually maintained and reinforced in cognition. Therefore, to find ways of reducing or at least controlling negative consequences of stereotyping, we need to understand them also at the level of their mental representation. However, the precise semantic features of a mental model at this abstract level of stereotype representation are not currently known. In this thesis, I attempt to address this issue by integrating the existing literature on stereotyping with the perspective of grounded cognition.

In the context of grounded cognition. Suitner, Maass, and Ronconi (2017) found that stereotypes about agency (males are perceived as more agentic than females; Maas, Suitner, Favaretto, & Cignacchi, 2009) are associated with the horizontal spatial dimension. They suggest that people hold associations between gender, agency, and *horizontal* positions in space. First, Suitner et al. (2017) demonstrated that participants associated males with agency, which in turn was associated with the rightward spatial trajectory (i.e., moving from left to right). Second, they found that in a reaction-time test, participants were more accurate to categorise pictures of male faces when the pictures of males showed their profiles facing the rightward trajectory (versus leftward), whilst the opposite was found for female faces. Such a pattern was possibly due to a congruency between males being seen as agentic, and their rightward-facing profiles. This pattern of results was reduced when participants were trained to associate female faces with the rightward trajectory, which also appeared to diminish self-reported benevolent sexism. This study indicates how cognitive associations formed on the basis of developed stereotypes are associated with sexist attitudes.

To conclude, it appears that gender stereotypes are unlikely to be based on actual biological differences between males and females. Given the negative consequences of stereotypes for how men and women are treated within the society, it is important to investigate them further, also on the level of their cognitive and representational features.

The significance of power in social cognition

In the current thesis, I focused on mental simulations associated with the concept of power. This is because power is an important psychological construct that impacts human cognition in multiple ways. For instance, Guinote, Willis, Martellotta (2010) found that feeling socially powerful increases implicit negative attitudes towards different races. In a series of studies, participants were led to believe that they either had high or low power (e.g., they thought they could or could not influence someone else's decision-making). Then, their racial implicit and explicit attitudes were assessed. Participant in the high-power, as opposed to the low-power condition, were significantly faster at categorising negatively valenced words after being primed with Black versus White faces. This implies that participants in the high-power condition associated Black faces with negativity. These findings were further replicated when assessing attitudes towards another racial group (i.e., Arabs) with an Implicit Association Test (Greenwald, McGhee, & Schwartz, 1998). Specifically, powerful participants had a higher tendency than powerless or neutral participants to associate Arabs with unpleasant words. Interestingly, manipulating the target group's power (i.e., high-power versus low-power Black person) did not affect racial prejudice. Overall, Guinote et al. (2010) demonstrated that a perceiver's high power increases implicit prejudice against stigmatised groups.

Similarly, Guinote and Phillips (2010) examined the effect of holding power on stereotyping. In their research, participants were recruited from a population of managers and subordinates who worked in hotels. Participants were presented with descriptions of potential candidates for a job. The descriptions were either stereotype-consistent or -inconsistent and included information about candidates' ethnicity (e.g., a stereotype-consistent description would present an English person applying for a teacher position, and a stereotype-inconsistent would present an Afro-Caribbean person applying for a teacher position). Both groups, managers and subordinates, were asked evaluate the candidates. It was found that managers (i.e., the powerful group) assessed out-group candidates by relying on stereotype consistency of the presented job. In contrast, subordinates (i.e., the powerless group) focused more on individuating information and did not rely on stereotyping. Such findings imply that power holders pay more attention to stereotypic information when it is relevant to a task at hand.

Furthermore, in a recent review of literature on power, Guinote (2017) concludes that power holders rely more on stereotypical beliefs, commit more self-serving biases, and they might be less socially attentive than those who are not powerful. Powerful individuals are focused on their own desires and needs to meet their goals. They are also motivated to maintain their power (e.g., Ratcliff & Vescio, 2013). Overall, the above findings indicate that power might be one of the most important factors influencing human cognition. Specifically, holding power increases biased attitudes and beliefs. Therefore, in my project, I focused on trying to better understand the representational features of power in relation to gender stereotypes.

Another set of findings suggesting that the concept of power is important in social cognition is that showing an automatic readiness to identify social hierarchies.

For example, Stewart et al. (2012) demonstrated that people are able detect dominance (an attribute associated with power) automatically at the preconscious level. These findings were corroborated by Rule, Adams, Ambady, and Freeman (2012). In their study, participants were asked to judge dominance from neutral, submissive, and dominant body poses. Despite time constraints, participants judged dominance levels based on poses more accurately than chance after seeing the stimuli for only 40ms. The time constraint did not impact the accuracy of their judgements. Overall, these studies indicate that people exhibit a readiness to process dominance cues. In line with the above studies, Zink et al. (2008) demonstrated that people process social hierarchies on the neural level even when hierarchy cues are irrelevant to their task performance. Participants were presented with a simple computer game requiring them to press buttons on a computer keyboard. The participants played the game simultaneously, ostensibly with other players. The other players' faces were presented on the computer screen. The social hierarchy or rank of the other players was manipulated by presenting different symbols of military rank next to their faces (e.g., three stars would indicate a high rank). Participants did not compete against the other players and the other players did not determine participants' success in the game. Nevertheless, the findings indicated that participants' brain activity (in the occipital/parietal cortex, ventral striatum, and parahippocampal cortex) was greater when viewing players with higher rather than lower ranks. Another study by Chiao et al. (2009) further demonstrated that people employ distinct neural pathways while representing social hierarchy, as opposed to number magnitudes. Overall, these findings suggest that people are able to detect social hierarchy cues, that is, infer targets' status and power, very efficiently, even following a brief exposure. These processes are reflected in distinct, status-related neural circuity.

Integrating grounded cognition, conceptual blending, and stereotyping

As noted above, the theory of grounded cognition assumes that thinking about abstract concepts implying sensorimotor properties involves a spatial simulation of these concepts (Barsalou, 2008). Within this process, response facilitation occurs when there is a congruency between a concept's physical spatial position, as perceived, and its spatial position in terms of its simulated mental representation. For instance, in the context of power, power is associated with upper vertical positions, so people are faster at selecting powerful groups when they are shown at the top of the visual field than at the bottom (Schubert, 2005). I propose that spatial vertical simulation can be also associated with thinking about gender, such that the simulation of gender can be based on the simulation of power. That is, the simulation of gender is based on blending between gender and power, and that a stereotypic link (i.e., a higher congruency) between these concepts might involve *more pronounced* spatial simulations (that is, when seeing males as relatively powerful and females as relatively powerless). As a result, reasoning about social categories such as gender in a power context, and thinking about hierarchies amongst *social groups* might involve representations employing spatial processing. The analysis of processes underlying such representations can help to better understand the cognitive basis of socially-relevant categorisations and provide a deeper insight into cognitive integration and production of meaning.

The suggestions on blending associated with stereotypical associations and physical space find indirect support in the research by Suitner et al. (2017) that was described earlier in the chapter. They demonstrated spatial bases of associations between agency and gender on the horizontal dimension and rightward trajectory. The rightward trajectory refers to performing an action from the left to the right direction (e.g., writing direction). Training people to associate females with spatial features associated mainly with males (e.g., rightward facing faces) involved a reduction in benevolent sexism. These findings provide a hint that stereotypic categorisations involve spatial associations via a blend between gender and agency, and as such, potentially, spatial simulations of gender.

In the context of gender, power, and physical space, my line of reasoning is as follows. First, I argue that although many social groups might not a priori be strongly associated with power (different from the materials used in Schubert's research), spatial simulations can still occur, due to conceptual blending. This would most likely happen in situations when a) thinking about social categories is power-related, b) the stimuli per se do not carry a clear power connotation, but c) such a connotation arises when thinking about the stimuli in stereotypic ways. In fact, during blending, power might act as the context in which spatial features of verticality are activated (Schubert, 2005). Consistent with this argument, as mentioned earlier, Lebois et al. (2015) claim that the spatial features of grounded concepts are not automatically activated in all contexts, but rather that spatial processing *only* takes place when the grounded features of mentally represented concepts are relevant to the conditions under which the processing occurs. That is, power would be more likely simulated in space when people think about concepts that are semantically related to power (e.g., judging which person - *master or servant* - is the powerful one, Schubert, 2005). However, it is unlikely that social groups are spatially represented when the processing does not occur in the context of power (e.g., judging which word - *master or servant* - has more letters; see also Taylor et al., 2015). Congruent with this perspective, I propose that the perception of social groups, such as gender, in terms of power (males-powerful/powerless; females-powerful/powerless) acts as a context in which the processing of information about gender is likely to be linked to the spatial features of power. In other words, I suggest that the vertical simulation of power, as demonstrated by Schubert (2005), will take place when people think about concepts in power-related terms. This could be either concepts that carry power implications by definition (e.g., Schubert's *master* versus *servant*) or, alternatively, concepts that are per definition not tied to power levels (e.g., *male* versus *female*), but may be applied to power when people need to differentiate power levels between gender, such that these two concepts are likely to be blended due to stereotyping.

Second, as gender-power associations are malleable and their blending is dependent on the context, I suggest that stereotype-consistent blending (males-powerful; females-powerless) is associated with spatial simulations. The socio-political structure within Western societies has been based primarily on patriarchy where males usually have higher social status and more economical/political power than females (Ellemers, 2018; Rudman & Phelan, 2010; Wood & Eagly, 2002). People are more likely to attribute stronger, socially powerful characteristics to males and weaker, powerless characteristics to females (Eagly, Makhijani, & Klonsky, 1992; Eagly & Steffens, 1984; Rudman, Moss-Racusin, Phelan & Nauts, 2012). Further, females who contradict the stereotype of having low status (i.e., when there is low stereotype-fit) tend to be negatively evaluated by both females and males as a result of the male-powerful link (Rudman & Kilianski, 2000). Interestingly, this latter research found that participants' attitudes towards high-authority females were more negative than attitudes towards low-authority females, and both low- and high-authority males. As applied to the research in this thesis, I suggest that in the absence of any other diagnostic information except for gender, spatial simulation is more likely to take place when a stereotypic view of a person or group is being processed and the stereotypic attribute of that person or group is associated with a concept grounded in space – in this case, power.

Also, because there are individual differences in gender-stereotypical beliefs (Miller & Saucier, 2018; Rudman & Kilianski, 2000), I investigated whether the extent to which a person links males with high power and females with low power is related to spatial simulations of males at the top and females at the bottom. Individuals who tend to more strongly associate males with power might also exhibit more pronounced spatial simulation of powerful-males at top versus females. A more accessible link between the simulated concepts, that is, gender and power, might provide more processing advantage leading to more pronounced simulations.

Integrating work on cognitive blending (Fauconnier & Turner, 1998), the context-dependency of spatial congruency effects (e.g., Lebois et al., 2015), the stereotypic link between males and power (Ellemers, 2018; Rudman & Kilianski, 2000) and connections between power and verticality (Schubert, 2005; von Hecker et al., 2013), I suggest that blending between *male* and *powerful* as well as between *female* and *powerless* (i.e., stereotype-fit) should possess a processing advantage and be associated with pronounced power-related spatial simulations due to the stereotypic associations. Subsequently, if stereotypic thinking and power context are essential in eliciting spatial simulations in experimental conditions, removing or reducing a possibility of making stereotypic judgments about gender in the context of power should not stimulate spatial processing. To clarify, by stereotypic thinking I refer to a general tendency to think according to gender stereotypes. When such tendency is applied to making judgements in the context of power, then people should be more likely to consider males as powerful and females as powerless.

To summarise, I believe that integrating existing literature on grounded cognition with research on stereotyping and consequences of power is essential, as the knowledge we currently possess in this respect is limited (e.g., Giessner & Schubert, 2007; Guinote, 2017; Meier & Robinson, 2004; Schneider, Rutjens, Jostmann, & Lakens, 2011; Schubert, 2005; Slepian, Weisbuch, Rule, and Ambady, 2011; Suitner et al. 2017). Most of the reported studies provide evidence for spatial processes in social thinking, however, they fail to explain the nature and significance of spatial thinking for social processes. Similarly, research focusing on stereotyping mainly addressed its consequences for discrimination and attitudes rather the representational structure (Fiske & Glick, 2001; Logel, et al., 2009; Rudamn & Kilianksi, 2000; Rudman & Phelan, 2010). Yet, understanding the representational structure of stereotypes is also essential in designing interventions against discrimination or prejudice. This is especially important in the context of social power, as social power seems to be an important factor in maintaining social inequality and promoting prejudice (Guinote et al. 2010; Guinote, 2017). Therefore, I attempted to address these issues in my thesis, in the hope of stimulating more research into the cognitive associations between space and abstract concepts in social thought. Below, I present an overview of the studies that were designed to accomplish the above aims.

Overview of studies

Overall, I conducted 14 studies. First, I designed six behavioural experiments using reaction-time methodology investigating the presence of spatial simulations of gender via power perception, which are described in Chapter 2. Subsequently, in Chapter 3, I present a follow-up experiment that was designed to replicate the previous findings by using a physiological measure of pupil size (pupillometry). This is because pupil dilation has been found to be a proxy of cognitive processing in terms of cognitive effort or load and expectancy violation (Aston-Jones & Cohen, 2005; Proulx, Sleegers, & Tritt, 2017; Smallwood et al., 2011).

In the next two chapters, I present studies examining the context-dependency of spatial simulations whereby task presentation was manipulated in terms of stimuli and instructions (Chapter 4) and stereotypic associations with regard to accessibility (Chapter 5). In my last empirical chapter (Chapter 6), I demonstrate the impact of vertical presentations of gender for stereotypic judgements and perception of size. Finally, in the concluding Chapter 7, I integrate and discuss the findings across all studies and provide an overall conclusion.

Spatial simulations and stereotypic thinking

First, I tested my main aim of exploring spatial simulations via social power perception and gender stereotypes in a linked series of six experiments (Chapter 2). In order to provide an overall picture of the investigated spatial effects, I conducted an overall analysis by integrating the data across all experiments and estimating a hierarchical linear model, in which I treated *participants* as random factors nested within a contextual variable - *studies*. Such an integration was possible, as I used the same basic experimental paradigm in all six studies. However, I slightly modified the procedure of individual experiments to assess the nature of the effects across different contexts. To ensure that participants thought about *social power* in my experiments, I provided its definition. Further, I tested whether stereotype-fit (males-powerful; females-powerless) would be associated with stronger spatial processing, such that powerful-males would be detected faster at the top in the visual field than powerful-females. Initially, I explored these ideas by asking British (Study 1) and Polish (Study 2) participants to quickly categorise either the powerful (condition 1) or the powerless (condition 2) person on the basis of gendered names only (male or female) that were vertically displayed (top or bottom) on a computer screen. Subsequently, in Study 3, I introduced more control over participants' associations by pairing both male and female names with high-status and gender-neutral professions. In this way, participants could rely not only on gendered names, but also on the additional information (i.e., professions) when making decisions about powerful/powerless persons. In Study 4, I explored whether thinking about power and gender in terms of social status was associated with more pronounced spatial representations of males at the top and females at the bottom. I tested this idea by priming participants with either social status items (e.g., scientist, professor, and dentist) or neutral words (e.g., potato, carrot, and cucumber). In the final two experiments of the chapter (Studies 5 and 6), I investigated the same assumptions as in Studies 1 and 2, but I aimed at directly comparing whether the response mode was associated with the extent to which spatial simulations occurred. That is, in Study 5 participants used vertical keys (arrows up-down) and in Study 6 horizontal keys (A-L) to indicate their responses. Finally, I examined potential correlates of the investigated spatial simulations by measuring participants' stereotypic and concrete associations: male-high status (Studies 1, 2, 3, and 4); male-rationality; top-rationality (Studies 5 and 6).

Across all experiments, I expected that participants would be significantly quicker at detecting male names as powerful when they appeared at the top as opposed to female names as powerful at the top. Similarly, I predicted faster responses to females as powerless when they appeared at the bottom as opposed to males as powerless in the same position. Finally, because the power concept per se is grounded in space (Schubert, 2005), I explored whether participants would be quicker to detect both males and females as powerful at the top versus bottom, and males and females as powerless at the bottom. This could be possible via an association between power and top positions and lack of power and bottom positions regardless of gender. However, if conceptual blending between power and gender via stereotypic associations is essential for spatial simulations to occur, then only males considered as powerful should be simulated at the top versus bottom and vice versa for females as powerless.

Measuring power-gender spatial simulations with pupillometry

In a follow-up study, I assessed power-gender spatial simulations with a physiological method involving pupillometry (Chapter 3). This was done to test the robustness of the findings obtained in my initial studies presented in Chapter 2. Pupil size has been found to be associated with expectancy violations and cognitive effort, such that pupil dilations reflect higher surprise (i.e., when unexpected stimuli are encountered) and cognitive load associated with increased cognitive effort (Proulx et al., 2017; van der Wel and van Steenbergen, 2018). As pupillometry represents an alternative and complementary way of investigating cognitive processing to the reaction-time methodology, I applied this method to investigate spatial grounding of power in relation to gender stereotypes. In Study 7, to test the validity and relevance of pupillometry to my project, first, I replicated Schubert's (2005) spatial task. I asked participants to quickly categorise powerful individuals on the screen while presenting them with vertically positioned (top versus bottom) power-related words (e.g., servant, master; adapted from Schubert, 2005) in one block. Second, in another block, I replicated my own task used in Study 1 of the present thesis. Specifically, I presented participants with vertically positioned gendered names and asked them to quickly categorise powerful persons. I predicted increased pupil dilation on trials that represented stimuli in *incongruent* spatial locations (e.g., female names merely presented at the top or categorised as powerful at the top, and powerless groups, e.g., servants, presented at the top). Pupil dilations on such trials would index surprise and higher cognitive load (see Proulx et al., 2017).

Dependency of spatial simulations on task features

Next, I present three further experiments that explored the context dependency of spatial simulations (Chapter 4). As power has been found to be grounded in space

(Schubert, 2005), I further investigated whether any individual that is associated with power (regardless of gender) would be spatially simulated (at the top or bottom) due to verticality cues (i.e., presenting names at the top or bottom). Therefore, in Studies 8 and 9, I presented participants with pairs of gendered names, as in Studies 1 - 6, but this time the pairs were both males or both females. That is, participants made judgements about the powerful or powerless person by distinguishing between two females or two males. In Study 8, I paired each gendered name with a high-status and gender-neutral profession (adapted from Study 3), whilst in Study 9, I presented only names. Given that power was always associated with space across contexts, I expected that whichever individual was associated with being *powerful* would be detected faster and simulated at the top, whilst any individual associated with being *powerless* would be identified quicker and simulated at the bottom.

In Study 10, I tested whether gender is simulated spatially when people do not think about males and females in the context of power. To do that, I presented participants with gendered names that were displayed vertically on the screen. Only one name (either male or female) was presented in each trial either at the top or bottom. I asked participants to quickly categorise the presented names as male or female. Given that gender per se was associated with spatial simulations, I predicted that male names would be categorised faster as male when presented at the top, whilst female names would be categorised faster when displayed at the bottom. However, if gender itself was not grounded in space, I predicted no differences in participants' response latencies to classify both genders at the top and bottom.

Dependency of spatial simulations on stereotype accessibility

To provide a complete picture of the contextual effects on spatial simulations, in Chapter 5, I present two studies in which I investigated whether manipulating stereotype accessibility could involve the presence of spatial processing in mental representations of gender and power. Therefore, in Study 11, I asked participants to categorise powerful individuals on the screen under restricted or relaxed task conditions to manipulate stereotype accessibility. In one condition, participants had only 600ms to respond, whilst in the other one 2000ms. The literature indicates that people attend to cues that are more immediately relevant and accessible to the task (Macrae, Bodenhausen, Milne, Thorn, & Castelli, 1997). Therefore, under very restricted conditions (600ms) it is less likely that participants would be able to use the stereotype to make their judgements. By contrast, it is likely that the verticality cues are easy to process (top versus bottom), because they do not require activating social knowledge (i.e., stereotypes). Because of this, the time constraint of 600ms might be sufficient to process and use those cues in making power judgements. In contrast, 2000ms should provide sufficient processing time to activate and apply stereotype-consistent knowledge in conjunction with the verticality cues. At the same time, because 2000ms is still quite a short period of time, under such conditions it is unlikely that participants would be able to effectively inhibit stereotypic associations on most of the trials. Hence, I predicted that in the restricted-time condition, participants would be significantly more likely to rely on verticality cues when making a judgement about powerful individuals, as the cues would be more salient. That is, participants would be more likely to pick powerful individuals at the top regardless of gender. In the non-restricted condition, I predicted the involvement of stereotypic thinking, as participants would be given more time to think about their judgements. Specifically, stereotypic thinking might involve more processing time than verticality cues. Therefore, participants should be more likely to rely on such biases rather than focus on more immediate spatial cues.

If stereotypic thinking was involved in spatial simulations (Rudman & Kilianski, 2000), I further predicted that priming people with powerful females and powerless males should reduce the spatial simulation effects of males as powerful at the top and strengthen the representation of females as powerful at the top – given that such simulations were previously detected. Therefore, in Study 12, I asked participants to complete the spatial task (the same as used in Studies 1 - 7), and quickly identify the powerful person on the screen on the basis of pairs of opposite gendered names (male - female). However, one group of participants was primed with powerful-females and powerless-males before completing the spatial task. I predicted that the spatial processing would be significantly reduced in the priming condition as opposed to the control condition.

The impact of spatial presentation on gender perceptions

Finally, in Chapter 6, my aim was to test whether simulations of gender on the vertical dimension in the context of power would bias participants' subsequent stereotypic judgments about males and females. In Study 13, I tested whether presenting male and female names at the top would be associated with a biased perception of their physical size. The grounded cognition perspective suggests that mental representations of abstract concepts (e.g., power) can be associated with perceptual experiences of both verticality and *size*. Metaphorically, *physical size* is associated with *being on top*, which then relates to greater power. To test whether participants would associate gender with an enhanced physical size after being primed with a vertical spatial location of males/females, I asked participants to attend to a vertical presentation of gendered names (in one condition male names appeared at the top and female ones at the bottom, whilst in the other condition this was reversed – females at the top and males at the bottom; or no presentation at all in the control condition). In the two experimental conditions, participants were also informed that the presented individuals (male and female names) represented managers and subordinates. Then, after each priming trial, participants were asked to estimate font sizes of subsequently centrally shown male or female names. I predicted that participants presented with male names at the top would judge subsequently presented male names on the computer screen as bigger in size (i.e., overestimate their actual sizes), whilst the female names would be underestimated.

In Study 14, I wanted to test whether spatial vertical simulation of males as powerful and females as powerless would affect participants' judgements about gender in terms of social dimensions associated with Stereotype Content Model, that is, competence and warmth (Fiske, Cuddy, Glick, & Xu, 2002). To investigate this issue, again I asked participants to complete a spatial task (as in Studies 1 – 6), but this time instead of names, I presented pictures of opposite gender pairs (one male; the other one female) that were matched in terms of age, masculinity/femininity, and attractiveness. Participants were asked to quickly categorise the powerful or the powerless person on the basis of the pictures. After the spatial task, I asked them to rate the competence, warmth, gender typicality, and how much they liked each person presented in the spatial task. I predicted that males as powerful simulated at the top would be judged as more competent and less warm in contrast to females as powerful or powerless presented in the same spatial location.

Overview of the analytical strategy

For each study that involved measurement of response latencies and pupil size, I estimated a linear mixed model. Before conducting each analysis, I assessed the appropriate random structure that would best fit my data and then I estimated a final linear mixed model to evaluate my fixed effects, that is, gender choice (males versus females) and trial type (males-top versus females-top; see Jaeger, 2008; Judd, Westfall, & Kenny, 2012). The way of determining the best random structure for each study is presented in Appendix 1.

To normalize the distribution of response times to the chosen categories, the reaction-times of each participant were first range restricted to 3000ms and subsequently trimmed using the Tukey criterion (response latencies that are larger or smaller than the upper or lower quartile, respectively, plus or minus 1.5 times the interquartile range for each participant are removed, see Clark-Carter, 2004). In this thesis, effect sizes pertinent to the target hypotheses are reported as Cohen's *d* for chi-square tests and *dz* for linear mixed models analyses.

The IAT scores were first computed by recoding values that were lower than 300ms as 300ms, and those higher than 3000ms as 3000ms (Rudman & Kilianski, 2000). Subsequently, I subjected the scores to logarithmic transformations to normalise the distribution. Finally, after this transformation, I computed mean response times for each category of interest (e.g., mean scores for participants' responses to males-high-status items and females-high-status items). The same procedure was applied to every IAT reported in this thesis.

Implications

To summarise, in my thesis, I present a series of experiments that were designed to understand the representational and sensorimotor features of gender stereotypes. Although substantial research has already been conducted in the area of grounded cognition and stereotyping separately, not many studies focused on integrating these two areas of research. I believe that combining knowledge about cognition and stereotyping might provide better understanding of both of these domains. That is, investigating mechanisms involved in mental representation of stereotypes might provide knowledge about the nature of stereotypes per se, but also the functioning of human mental processes. Accumulating such knowledge could provide more effective ways of dealing with socially undesirable cognitive biases.

Chapter 2: Spatial simulations in power-gender associations (Studies 1 – 6)

Overview

In this chapter, I present six studies that were designed to test whether spatial simulations are involved in mental representations of gender and power on the vertical dimension. Overall, I used the same experimental paradigm across all studies, but each of the designs differed slightly to assess generalizability. In the main paradigm, participants completed a spatial task and categorised as quickly as possible which person (one female name; one male name) of two vertically displayed on the screen (top and bottom) was powerful (condition 1) or powerless (condition 2). Their choices of male and female names judged as powerful or powerless in each vertical position were recorded. I also measured participants' response latencies to their judgements. Finally, participants were asked to complete a standardised Implicit Association Test measuring their attitudes towards social status of males and females. In line with my hypothesis, power-gender stereotypic associations involved simulations of males as powerful at the top as opposed to (a) males as powerful at the bottom as well as (b) females as powerful at the top. Contrary to expectations, no spatial effects were detected when participants selected females as powerful and both male and female names as powerless. The effects were independent of implicit attitudes towards gender.

In a series of six experiments presented in this chapter, I tested whether thinking about gender in terms of power was associated with spatial simulations (males and females as powerful simulated at the top; males and females as powerless simulated at the bottom) and especially pronounced simulations in the case of stereotypic-thinking (higher proportions of choices of and faster responses to stereotype-consistent stimuli, i.e., males as powerful at the top/females as powerless at the bottom).

In terms of the analytical strategy, I followed suggestions by Lakens and Etz (2017). In their recent work regarding multiple study lines of research, the authors used likelihood ratios to demonstrate that when testing the same hypothesis in several experiments, there is a high probability of obtaining non-significant results in some of the experiments. In order to assess whether a hypothesis is indeed supported across several studies, there is a need for conducting an integrative analysis. Because of this, I present a combined analysis across the studies, but the results of each individual study are available in Appendix 6. As some of the individual studies did not achieve standard levels of statistical significance, to test my hypotheses regarding spatial simulation of gender and power on the vertical dimension, I estimated a hierarchical linear model that integrated the data across all my experiments. Below, I present the design of each of the studies.

Studies 1 & 2: Exploring spatial simulations in stereotypic thinking

Studies 1 and 2 were exploratory. Within both studies I used a similar reaction-time spatial task as Schubert (2005), but instead of powerful/powerless groups, I presented participants with pairs of male and female names as stimuli (e.g., Emily-Oliver) vertically displayed on a computer screen (one at the top; one at the bottom). Participants indicated as fast as possible which person was powerful (condition 1) or powerless (condition 2) on the basis of the names with horizontally arranged keys on the computer keyboard. Half of the participants pressed A to choose powerful and L to choose powerless persons, or vice versa for the other half of participants. In this task, I measured how often participants picked each name and I also recorded their response latencies to all the choices they made. I would like to note that the experiments were exploratory and therefore I aimed at allowing participants to freely make their own choices about powerful and powerless individuals. Specifically, contrary to Schubert's

(2005) design, in my experiments participants could not make a correct or incorrect judgement. As they were presented with gendered names only, rather than with generic powerful/powerless groups as in Schubert's research, they needed to decide themselves which person was the powerful/powerless. That is, I used a quasi-experimental design whereby the choice of the powerful/powerless person (either the male or female name) was the self-generated quasi-experimental variable.

Subsequently, participants completed an Implicit Association Test (IAT; Greenwald et al., 1998) measuring their implicit attitudes towards gender with regards to status. In a standardised IAT, participants are presented with items (one at a time in the middle of the screen) that belong to different categories (here, high or lower social status and male or female names). Participants are asked to quickly sort items to different categories. In the social-status IAT, high-status items and male names need to be categorised with one keyboard key and low-status ones and female names with another within one block. The key arrangement is then reversed in another block (one key for high-status items and female names and another for low-status items and male names). Then, participants' reaction-times are measured across these blocks. Faster response latencies to males-high-status items versus females-high-status ones indicate an implicit bias against females who have a high social status.

In Study 1, I recruited a sample of English-speaking participants, and in Study 2, Polish-speaking participants, to assess generalizability

Study 3: Controlled gender-power associations

In Studies 1 and 2, participants made their judgments about power on the basis of gendered names only. In the absence of any additional information, they were likely to use prior knowledge and freely accessible associations about social power of males and females. In Study 3, I therefore introduced more control over the thought context, in the hope to replicate the findings from Studies 1 and 2 with a small alteration to the method. Specifically, I aimed at creating experimental conditions that would be less ambiguous and enhance participants' representations of the processed concepts (i.e., power and gender). Specifically, I constrained the context for making the decision about powerful/powerless individuals by pairing each gendered name with a high-status profession in the spatial task in order to stimulate thinking about power implications. In addition, I asked participants to use vertical keys on the computer keyboard, to reduce potential cognitive load in relating the horizontal keys to the vertically presented stimuli on the screen. I hypothesised that constraining the context would involve spatial simulations of both males as powerful and females at the top (as power should be associated with verticality in general, see Schubert 2005), but such simulations would be more pronounced in the case males as powerful rather than females, as a result of stereotype consistency. I derived the predictions from Lebois et al.'s (2015) research, as it seems that introducing more constrained/relevant context to the thought processes would enhance the salience of the spatial features of grounded concepts (i.e., power) and vividness of the mental representation (Nisbett & Ross, 1980).

Study 4: Priming social status

Study 4 tested whether thinking about social status interacts with thinking about power of males and females to produce even stronger spatial simulations. In this study, I adapted the same design as in Study 1. However, this time before each trial of the spatial task, participants were primed with either social status items or neutral words. I expected that priming with social status would involve more pronounced spatial simulations when thinking about both males and females, especially in the case of thinking about males in stereotypic ways (males-powerful). To further examine if the implicit associations between gender and high status are related to spatial simulation, I asked participants to complete a social-status and gender IAT (adapted from Study 1).

Studies 5 and 6: Correlates of spatial simulation

In Studies 5 and 6, the main aim was to conduct a replication of Study 1 and to explore whether individual differences in stereotypical beliefs about gender (other than associations between genders and social status) may relate to spatial simulations of males at the top and females at the bottom. Therefore, I propose that there are additional conceptual associations that may support spatial simulation of the gender-power hierarchy.

Logically, a potential interacting domain would need to be metaphorically associated with vertical positioning and gender. Interestingly, Cian, Krishna and Schwarz (2015) reported that people hold associations between rationality and top positions as well as emotionality and bottom positions. Furthermore, it seems that females are stereotypically considered to express more emotions than males (e.g., see Glick & Fiske, 1996; Plant, Hyde, Keltner & Devine, 2006). I therefore hypothesised that links between rationality/emotionality and gender as well as rationality/emotionality and vertical positions might moderate or be related to the spatial simulation of gender on the vertical dimension similarly.

The design of two final studies of this chapter was based on Study 1. Across both experiments, I asked participants to complete the same spatial task as before, in which I presented vertically arranged gendered names. The only exception in the methodology was that in Study 5, I facilitated participants’ responses by asking them to use arrows up/down to indicate the powerful/powerless individuals on the screen, as in Study 3. In Study 6, participants responded in the usual way by using horizontally arranged keys, as in Study 1. After the spatial task, I asked participants to complete two

IATs. The first one measured associations between rationality/emotionality and gender in order to assess whether participants indeed held associations between females-emotionality and males-rationality (Plant et al., 2006). The second one was a replication of the IAT used by Cian et al. (2015) to support the assumption that people associate emotionality with bottom vertical positions and rationality with top ones.

Overall predictions across six studies

First, in terms of proportions of gendered names chosen as powerful and powerless, I predicted that participants would be more likely to pick males as powerful than females as powerful, due to stereotypic associations. In terms of reaction times, I hypothesised that if stereotypic thinking about gender in terms of power involved spatial simulation (Giessner & Schubert, 2007; Lebois et al., 2015; Schubert, 2005), then responses should be faster to male names appearing at the top than female names appearing at the top, for trials in which participants were asked to select who was powerful. Likewise, responses should be faster to female names appearing at the bottom than male names in the same location, for trials in which participants were asked to select who was powerless. However, to consider such effects as associated with spatial simulations and particular spatial locations, I also predicted that overall, the so-identified “powerful” name would be responded to quicker when at the top than at the bottom, and vice versa for the “powerless” name (see Schubert, 2005). However, if stereotypic associations were crucial in blending among gender, power, and space, then when participants selected females as powerful, no response facilitation should occur when females appeared at the top versus bottom. The same should be true for choices of males as powerless when their names appeared at the bottom versus top.

Finally, I predicted that holding male-high status, male-rationality, and top-rationality associations would likely involve quicker responses in selecting males as powerful at the top and females as powerless at the bottom (Rudman & Kilianski, 2000). This was tested by asking participants to complete a series of IATs, in which they were instructed to quickly sort items that belonged to different categories, e.g., quicker responses to sorting male names and high status labels with one key rather than females and high status would indicate an association between males and high status.

Method[4]

Participants

Demographic information about participants is presented in Table 1. The sample size of individual studies was informed by relevant previous research (e.g., Schubert, 2005; von Hecker et al., 2013, for medium effect size, Cohen's $dz = .25$). Participants for five of the studies were recruited from a sample of Cardiff University students and received either course credit or payment for their participation. Participants for the remaining study were recruited from Maria Curie-Sklodowska University of Lublin, Poland, and received course credit. Across six experiments, I recruited 372 participants in total.[5]

Exclusions: One participant was excluded from Study 1 and Study 3 due to extreme response times, as determined by Tukey criterion (specifically, their response times were larger than the upper quartile plus 1.5 times the interquartile range). In other words, their average response latencies were 2 standard deviations above the grand mean of all participants).

[4] Method section of each individual study is available in Appendix 2.

[5] The high proportion of female participants in my studies is attributable to the demographic makeup of the participant panels.

Table 1. Demographic information about participants.

Studies	N	Gender	Mean Age	Nationality
Study 1	79	66 females	20.17 years	British
Study 2	73	58 females	20.19 years	Polish
Study 3	52	47 females	19.04 years	British
Study 4	58	51 females	18.97 years	British
Study 5	59	57 females	22.12 years	British[6]
Study 6	60	52 females	21.40 years	British

Materials

All Studies. Spatial Task. I selected 10 popular British (Studies 1, 3, 4, 5, and 6) and Polish (Study 2) names (5 female and 5 male).[7] Each name was randomly paired with a name of the opposite gender (e.g., Oliver – Emily, see Appendix 3 for all the pairs). Across all studies, the matched pairs were then presented on the computer screen in white letters on a black screen (font size 15 – 21 across the studies). In Study 3, I additionally selected professions (derived from a pilot study)[8] that were assigned to male and female names (adapted from Study 1; e.g., Oliver-Professor, Sophie-Professor). Each pair (profession and name) was assigned to a corresponding pair including a name of the opposite gender. I created all combinations of professions and

[6] 26% of the sample were non-native English speakers, but were fluent in English.
[7] See the most popular names: Popular British baby names: Year by year. (n.d.). Retrieved February 9, 2015, from http://www.babycentre.- co.uk/popular- baby-names (Study 1a); Academy of childbirth. (n.d.). Retrieved May 2, 2016, from: http://akademiaporodu.pl/top-news/najpopularniejsze-imiona-w-2015-2016-mapy-ranking (Study 2).
[8] In a pilot study, 30 participants rated 30 professions on levels of power, agency, gender-typicality and social status (0: low; 10: high). The mean ratings revealed five professions (doctor, dentist, architect, scientist, professor) with high power (M = 6.92; SD = .85), social status (M = 8.20; SD = .83), being gender-neutral (M = 4.34; SD = .53) and agency-neutral (M = 4.68; SD = .83).

gender (e.g., male scientist - female professor; female professor - male scientist) within four sets of pairings (see Appendix 4 for all combinations). Participants were randomly assigned to receive one set of combinations. In Study 4, I introduced another minor modification. Before each trial, (in which they were required to indicated the powerful/powerless person at the top or at the bottom) participants were presented with a social status item (gender-neutral profession adapted from Study 3, i.e.: scientist, architect, doctor, professor, dentist) or neutral word (i.e., vegetables: carrot, potato, lettuce, broccoli, cabbage). The prime was manipulated within-participants. After making decisions about the powerful/powerless person in each trial, participants were asked to report whether the initially presented prime word belonged to the social status category (by pressing the arrow pointing left), or to the vegetable category (by pressing the arrow pointing right).[9] In this way, the prime word was activated during participants' decision about the powerful/powerless person.

Design

All Studies. Spatial Task. The same overarching design was used in all studies. Task or instructions ("find powerful" versus "find powerless") was always manipulated between-participants while the trial type (males-top and females-bottom versus females-top and males-bottom) was always manipulated within-participants. There were four blocks of trials. Each block included a presentation of five pairs shown twice (ten trials in total). In half of the trials, a male name (e.g., Oliver) was displayed at the top of the screen with a female name (e.g., Emily) at the bottom (see Figure 1). Both names were centred and there was a 23cm vertical distance between the names. In the remaining trials, this display was reversed (females at the top and males at the bottom). There were 40 trials in total. The order of trials within each block was

[9] The responses were reversed for the other half of our participants: they pressed the arrow pointing right for social status items and the arrow pointing left for vegetables (participants were randomly assigned to one of the key arrangements).

randomised. The same design was applied to Study 3, except that instead of presenting pairs of gendered names, I presented gendered names paired with professions (see Figure 2). In Study 4, prior to presenting the first half of the trials (males-top; females-bottom) participants were primed with a social status item, and prior to presenting the other half (males-top; females-bottom) they were shown a neutral word (the prime words were selected at random from a list of vegetables; see Figure 3). The same was done for the trials where females were presented at the top and males at the bottom, so there were 20 trials within each block (80 trials in total). The trials were also randomised within each block within-participants; there were four blocks in total. I measured how quickly participants responded and which gender they picked as powerful/powerless on each trial type (see Table 2 for an overview of the manipulated variables).

Table 2. Overview of the manipulated factors.

Factor	Levels	Design
Task	Powerful versus powerless	Between-participants
Trial type	Males-top versus females-top	Within-participants
Gender choice	Males versus females	Within-participants/quasi-experimental variable

Figure 1. Spatial task procedure of a single trial (Studies 1, 2, 5 and 6).[10]

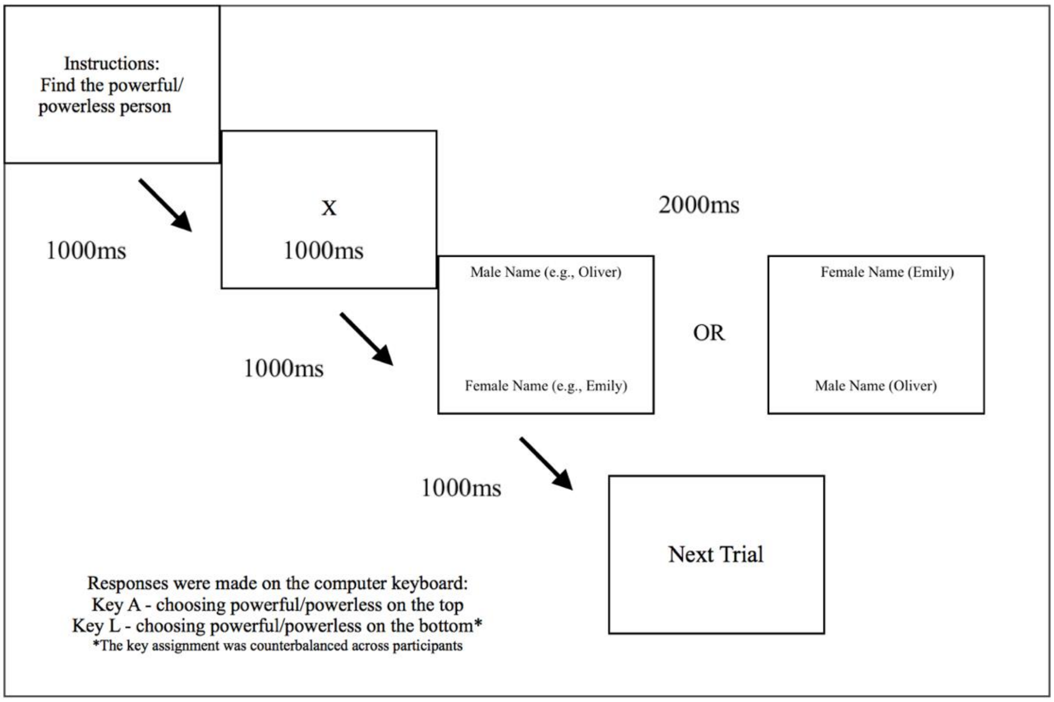
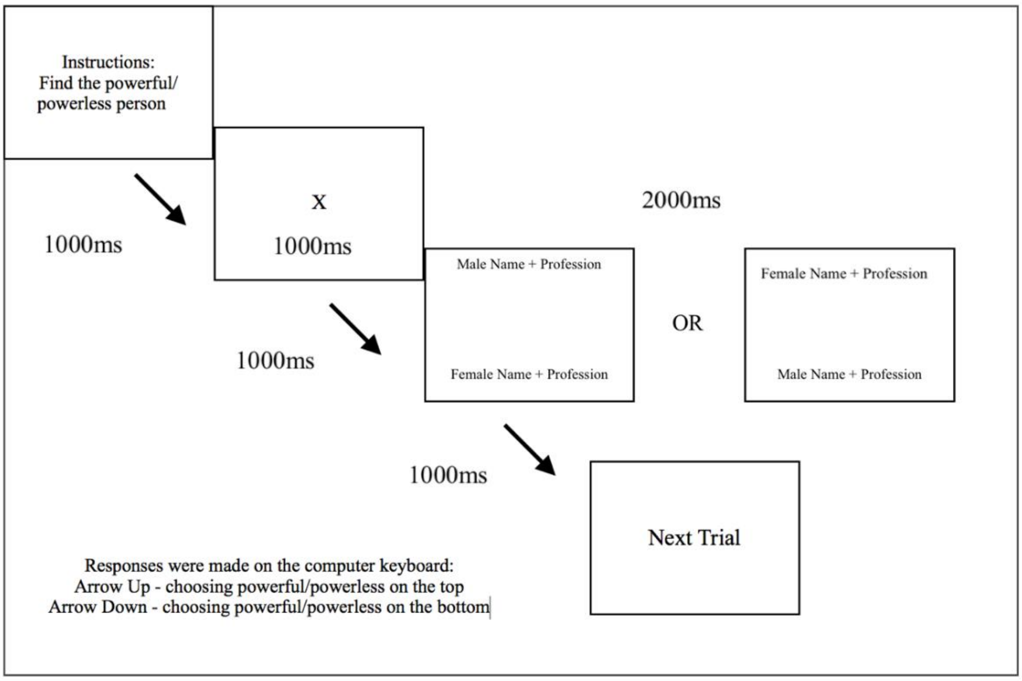
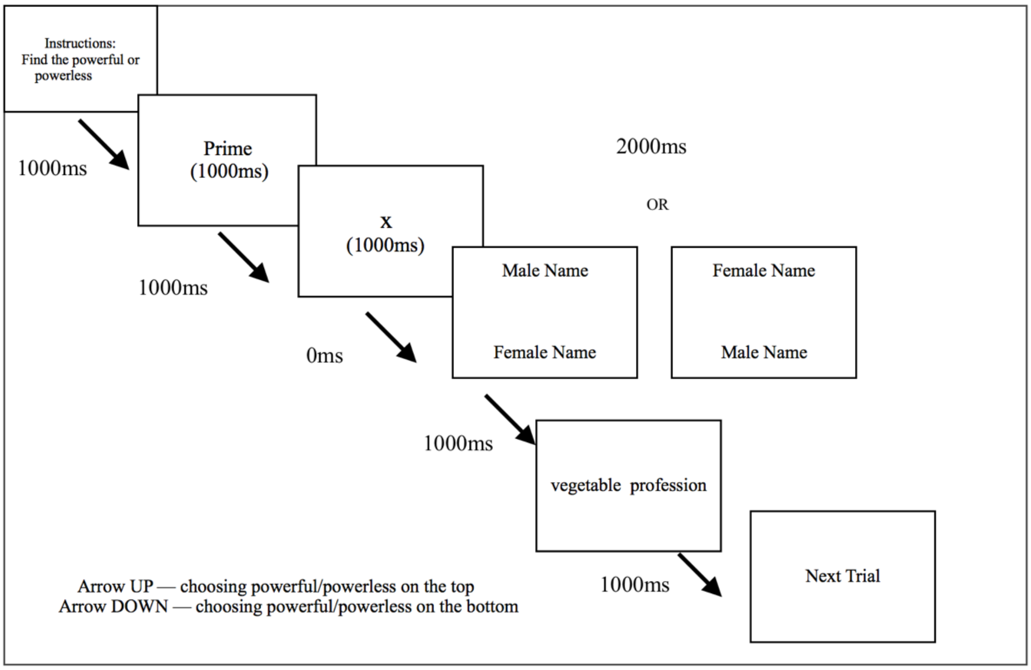


Figure 2. Spatial task procedure of a single trial (Study 3).



[10] In Study 5, participants used the arrow pointing up to indicate the powerful/powerless at the top and the arrow pointing up to indicate the target at the bottom.

Figure 3. Spatial task procedure of a single trial (Study 4).



Studies 1, 2, 3, and 4. Status-gender IAT. I adapted the standard IAT of implicit attitudes towards high- or low-status males and females from research by Rudman and Kilianski (2000; see Appendix 5 for all the stimuli used). Within a single block, participants were asked to categorise items that belonged to either high status (boss, supervisor, expert, leader, executive, authority) and males (e.g., Paul, Brian, Kevin) or low status (secretary, helper, aide, clerk, subordinate, assistant) and females (e.g., Lauren, Kate, Zoe), or vice versa in another block.[11] The order of the blocks was counterbalanced across participants. In Study 2, the instructions and categories were presented in Polish. In Study 3, I reduced the number of blocks of a standard IAT from five to two (see Sriram & Greenwald, 2009). Participants were asked to keep in mind two target categories within two blocks (male and high status in one block or female and high status in another) and respond as fast as possible by pressing a designated key when they saw an item that belonged to the target category on the screen. When they

[11] In the current thesis, I disregard conceptual independence between power and status, and I focus on the conceptual overlap (see von Hecker et al., 2013).

saw distractors (female and low status or males and low status) they were to press a second designated key.

Studies 5 and 6. Rationality/emotionality-gender IAT. I asked participants to quickly and accurately sort items that belonged to the category of emotionality (feeling, mood, sentiment, sympathy) and females (e.g., Karen, Caitlin, Jasmine), as well as rationality (intelligent, logic, reason, thinking) and males (e.g., Brian, Kevin, Paul), and vice versa.

Studies 5 and 6. Rationality/emotionality-verticality IAT. Participants categorised items that belonged to the category of up (above, top, over, upper) and rationality, and the category of down (below, under, bottom, lower) and emotionality, or up/emotionality and down/rationality (the items for emotionality/rationality categories were the same as in the rationality/emotionality and gender IAT). Both IATs in Studies 5 and 6 were designed in the same standard way.

All Studies. Procedure

All studies were presented using DirectRT (Jarvis, 2012). Participants sat approximately 70cm away from the computer screen. After providing consent, participants read the instructions and received verbal definitions of *socially powerful* and *socially powerless* individuals (adapted from Galinsky, Gruenfeld, & Magee, 2003). In the powerful condition, I provided the following definition: ‘By powerful, we mean a person that controls the ability of another person to get something they want, or is in a position to evaluate that person’, whilst in the powerless condition I stated: ‘By powerless, we mean a person whose ability to get something they want is controlled by another person, and the other person is in a position to evaluate them.’ In Study 2, as there are no direct semantic equivalents of the adjectives *powerful* and *powerless* in Polish, I used substitutes. As social power can be defined in terms of social influence (Lewin, 1941), I asked Polish participants to find *influential* or *not influential* individuals on the screen. Then, participants completed the spatial task by pressing keys on the computer keyboard to indicate the powerful/powerless person on the screen. After the spatial task, all participants completed the IATs (the status-gender IAT in Studies 1, 2, 3, and 4; rationality/emotionality-gender and –verticality IATs in Studies 5 and 6). At the end of each session participants were debriefed. Each experiment lasted approximately 20 minutes.

Results

Proportions of choices. First, I analysed the proportions of participants' choices. As predicted, participants were more likely to choose males as powerful (59.5%) rather than females as powerful (40.5%), $\chi^2(1) = 658.78$, $p < .001$, $d > 2.0$, odds ratio $_{\text{choosing males as powerful versus females}}$ = 1.48 (95%CI[1.43, 1.53]. I also tested whether proportions of choices were correlated with reaction-times, but the analyses indicated that these two processes were independent, *ps* > .13.

Response latencies. In terms of reaction times of participants' choices, my data were hierarchical, so I built a multilevel linear model in which intercepts varied across *participants* that were nested within *studies.* Next, assessing the appropriate random structure that would best fit my data (see Appendix 1), I estimated the final model to evaluate the fixed effects, that is, gender choice (males versus females), trial type (males-top versus females-top), and task instructions (look for powerful versus powerless).

Overall, I found that the fixed effect of trial type was marginally significant, $F(1, 15825) = 3.15$, $p = .08$, $dz = .13$. Participants were marginally faster at responding when presented with males-top trials (M = 1060ms; SE = 17.52, 95% CI[1059, 1127]) than females-top (M = 1096ms; SE = 17.44, 95%CI[1068, 1137]. The fixed effect of gender choice (males: M = 1071ms; SE = 17.24, 95%CI[1057, 1125]; females: M =

1088ms; *SE* = 18.12, 95%CI[1069, 1140]) was also marginally significant, with quicker responses when choosing males versus females, *F*(1, 333) = 2.90, *p* = .09, *dz* = .04. However, the main effect of task was not significant, *F*(1, 405) = 1.56, *p* = .21, *dz* = .06, such that participants were equally fast to respond within the powerful (*M* = 1065ms; *SE* = 24, 95% CI[1030, 1124]) and powerless condition (*M* = 1093ms; *SE* = 23.81, 95% CI[1072, 1165]).

Next, I found that the trial type by gender choice interaction was significant, *F*(1, 16113) = 11.00, *p* < .001. Overall, participants were significantly faster to select males when they were presented at the top (*M* = 1055ms; *SE* = 17.65, 95% CI[1043, 1112]) than when they were presented at the bottom (*M* = 1088ms; *SE* = 17.64, 95% CI[1071, 1140), *t*(333) = -3.75, *p* < .001, *dz* = .14. Participants were also marginally faster at selecting males than they were presented at the top compared to selecting females when they were presented at the top, (*M* = 1104ms; *SE* = 18.42, 95% CI[1064, 1136), *t*(333) = -2.50, *p* = .06, *dz* = .12. However, other comparisons (females selected when they appeared at the bottom in contrast to top as well as males at the bottom) were not significant, *ps* > .71.

Further, the two-way interaction between gender choice and task was also significant, *F*(1, 340) = 15.74, *p* < .001. Overall, participants responded faster when selecting males as powerful (*M* = 1044; *SE* = 10.91, 95% CI[1010, 1102) as opposed to selecting females as powerful (*M* = 1100ms; *SE* = 10.69, 95% CI[1049, 1149), *t*(341) = -3.97, *p* < .001, *dz* =.20. Next, although participants were faster to select females as powerless *(M* = 1081ms; *SE* = 24.83, 95% CI[1061, 1159) compared to males (*M* = 1111ms; *SE* = 24.05, 95% CI[1080, 1174), this difference was not significant, *t*(331) = 1.61, *p* = .11, *dz* = .10.

Finally, as predicted by the conceptual blending account, the three-way interaction among gender choice, trial type, and task was significant, *F*(1, 16113) =

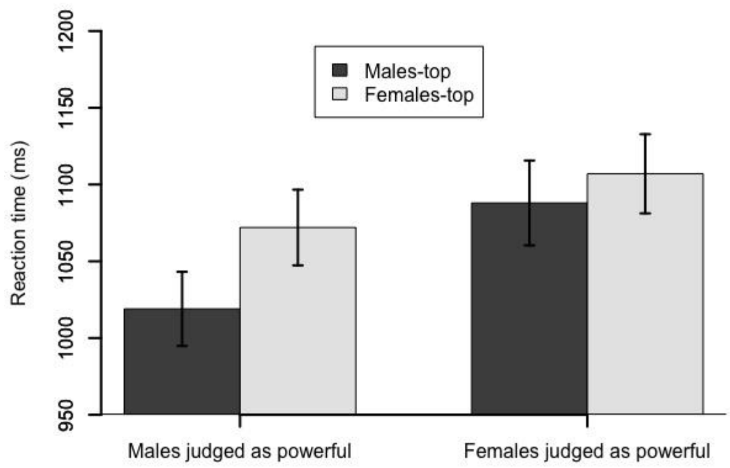
5.81, $p < .02$ (see Figure 4). Starting with the powerful condition, as hypothesised, I found that participants were significantly faster at selecting males as powerful when they were presented at the top of the screen (*M* = 1019ms; *SE* = 24.16, 95% CI[986, 1081) compared to selecting females as powerful when they were at the top of the screen (*M* = 1107ms; *SE* = 25.78, 95% CI[1038, 1140]), $t(597) = -4.38$, $p < .001$, dz = .20. Interestingly, in the powerless condition, participants did not differ in how quickly they selected males (*M* = 1108ms; *SE* = 24.66, 95% CI[1085, 1182]) versus females (*M* = 1057ms; *SE* = 25.51, 95% CI[1058, 1159]) as powerless when they appeared at the bottom, although the pattern of responses was consistent with my hypothesis, $t(733) = .85$, $p = .59$, dz =.13. Importantly, the contextual variable, *studies*, did not moderate the three-way interaction, $F(5, 16049) = 1.19$, $p = .31$. The above analyses indicate that when power (but not the lack of power) was conceptually combined with gender in a stereotype-consistent manner, participants spatially simulated gender on the top. However, this was not the case when a female was selected as powerful (i.e., in the context of a stereotype-inconsistent choice).

Subsequently, to determine whether the above effects were specifically associated with spatial locations, I conducted additional post-hoc analyses. These analyses revealed that indeed males were selected as powerful faster when they appeared at the top of the screen (*M* = 1019ms; *SE* = 9.34, 95% CI[991, 1088]) versus the bottom of the screen (*M* = 1072; *SE* = 24.65, 95% CI[1035, 1132]), $t(9530) = -4.72$, $p < .001$, $dz = .32$. The reverse pattern of results was true for choices of females, they were selected as powerful slower when they appeared at the top of the screen (*M* = 1107ms; *SE* = 27.04, 95% CI[1049, 1156]) versus the bottom of the screen (*M* = 1088ms; *SE* = 27.63, 95% CI[1070, 1178]), though this effect was not significant, $t(6329) = 1.66$, $p = .19$, $dz = .05$.

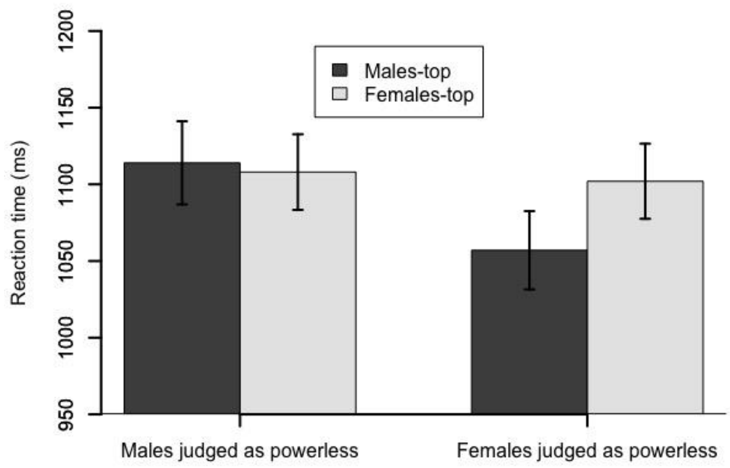
In contrast to my hypothesis, in the powerless condition, female names were not selected significantly slower when they appeared at the top of the screen (*M* = 1102ms; *SE* = 24.48, 95% CI[1072, 1168]) versus the bottom, (*M* = 1057ms; *SE* = 24.77, 95% CI[1067, 1165]), $t(9648) = -.41$, $p = .69$, $dz = .16$. This difference was also not significant when males were selected as powerless when they appeared at the top of the screen (*M* = 1114ms; *SE* = 27.11, 95% CI[1081, 1187]) versus the bottom (*M* = 1108ms; *SE* = 26.98, 95% CI[1098, 1204]), $t(6298) = -1.36$, $p = .19$, $dz = .02$. Overall, the three-way interaction suggests that the top-advantage in reaction times when choosing the powerful person is present in stereotype-consistent responding.

Figure 4. Response latencies to (a) males and females judged as powerful, and (b) males and females judged as powerless according to the trial type (males-top/females-bottom or females-top/males-bottom) across six experiments. The error bars show -/+1 standard error.[12]

(a) Powerful condition



(b) Powerless condition



[12] Note that "Males-top" refers to trials where male names were presented at the top and female names at the bottom, whilst "Females-top" represents trials where female names were presented at the top and male names at the bottom.

High status-gender IAT. To test whether the above findings were associated with individual differences in implicit associations between gender and social status, I analysed participants' IAT responses. After controlling for the relevant studies (1- 4), I found that participants were significantly faster at categorising males and high status items (*M* = 794ms; *SE* = 12.54, 95% CI[769, 818]) compared to females and high status items (*M* = 882ms; *SE* = 10.52, 95% CI[799, 841), $F(1, 256) = 5.25$, $p < .02$, $\eta^2_p = .02$. The studies by IAT category (males-high status; females-high status) interaction was not significant, indicating that the direction of the effect of the category did not differ across studies, $F(1, 256) = 2.54$, $p = .11$, $\eta^2_p = .01$. Second, I calculated a facilitation score for each participant. This was done by subtracting RTs from blocks in which participants responded to the category combination of males/high status from blocks in which they responded to females/high status; positive scores indicated a tendency to associate males with high status. Linking this index with my RT data, I also computed two separate difference scores subtracting the RTs for choices of males at the top from the RTs for choices of males at the bottom and, vice versa, the RTs for choices of females at the bottom from the RTs for choices of females at the top. The more positive these scores are, the more impactful is the simulation of males as powerful at the top. Therefore, I term these scores "simulation scores". After controlling for studies, I found that the IAT facilitation scores (i.e., the tendency to associate males with high status)[13] did not correlate with the spatial simulation of males as powerful at the top, $r(128) = -.07$, $p = .43$. The same was true for the simulation of females as powerless at the bottom, $r(104) = .06$, $p = .52$. Likewise, the correlation between facilitation score to males as powerful at the top versus females as powerful in the same spatial location and IAT facilitation score was not significant, $r(120) = -.06$, $p = .50$.

[13] I also correlated the absolute scores, i.e., participants' reaction times to choices of males as powerful on the top and females as powerless on the bottom, with the IATs facilitation scores, and I found similar results as when correlating the simulation scores.

Rationality/emotionality-gender IAT. After controlling for Studies 5 and 6, I found that participants had a significantly higher tendency to associate males with rationality (*M* = 697ms; *SE* = 9.41, 95% CI[678, 865) compared to females (*M* = 840ms; *SE* = 12.48, 95% CI[816, 865), $F(1, 117) = 4.25$, $p < .05$, $\eta^2_p = .04$. The choice category by studies interaction was not significant, $F(1, 117) = .56$, $p = .46$, $\eta^2_p = .01$. Similarly to the status-gender IAT, I calculated facilitations scores – the higher the score, the higher the tendency to associated males with rationality. Such associations did not significantly correlate with spatial simulations scores of males as powerful at the top, $r(58) = .11$, $p = .43$, or females as powerless at the bottom versus top, $r(57) = .18$, $p = .18$. The same was true for the correlation between the facilitation score to males as powerful at the top versus females as powerful at the top and IAT facilitation score, $r(51) = .21$, $p = .15$.

Rationality/emotionality-verticality IAT. I found that participants did not associate top vertical positions more with rationality (*M* = 784ms; *SE* = 13.68, 95% CI[757, 812) than with emotionality (*M* = 842ms; *SE* = 15.48, 95% CI[811, 873), $F(1, 114) = 1.50$, $p = .22$, $\eta^2_p = .01$ (after controlling for Studies 5 and 6). The interaction between the IAT category and studies was also not significant, $F(1, 114) = .97$, $p = .33$, $\eta^2_p = .01$. Subsequently, I calculated an IAT facilitation score – the higher the score the higher the tendency to associate rationality with top positions. The associations did not correlate with the spatial simulation of males as powerful at the top versus bottom, $r(56) = .11$, $p = .43$, or of females as powerless at the bottom versus the top, $r(51) = -.13$, $p = .35$. However, the facilitation score to males as powerful versus females as powerful at the top positively correlated with the IAT facilitation scores, $r(49) = .33$, $p < .02$, suggesting that participants who were more likely to associate rationality with top vertical position, were faster at detecting males as powerful at the top as opposed to females as powerful in the same position

Table 3. Summary of the aims and findings of each individual study.

Study	Aim	Findings
Study 1	To explore whether people spatially represent males as powerful at the top and females as powerless at the bottom when making judgements about gendered names.	1) Powerful-males-top (969ms; 95% CI[848, 1090]) vs. powerful-males-bottom (1040ms; 95% CI[904, 1176]), $p < .02$. 2) Powerful-males-top vs. powerful-females-top (1086ms; 95% CI[948, 1225]), $p < .01$. 3) Powerless-females-bottom (998ms; 95% CI[879, 1116]) vs. powerless-females-top (998ms; 95% CI[866, 1131]), $p = 1$. 4) Powerless-females-bottom vs. powerless-males-bottom (1017ms; 95% CI[883, 1151]), $p = .88$.
Study 2	To explore the same processes as in Study 1, however, by using a sample of Polish as opposed to British students.	1) Powerful-males-top (1143ms; 95% CI[1031, 1255]) vs. powerful-males-bottom (1194ms; 95% CI[1081, 1307]), $p = .13$. 2) Powerful-males-top vs. powerful-females-top (1218ms; 95% CI[1106, 1330]), $p < .01$. 3) Powerless-females-bottom (1208ms; 95% CI[1087.021 1330.375]) vs. powerless-females-top (1192ms; 95% CI[1080, 1305]), $p = .96$. 4) Powerless-females-bottom vs. powerless-males-bottom (1224ms; 95% CI[1108, 1339]), $p = .97$.
Study 3	To determine whether the spatial representation of gender would be more pronounced when people think about gender in the context of social-status cues.	1) Powerful-males-top (1175ms; 95% CI[1055, 1256]), vs. powerful-males-bottom (1175ms; 95% CI[1055, 1256]), $p < .01$. 2) Powerful-males-top vs. powerful-females-top (1108ms; 95% CI[996, 1221]), $p = .71$. 3) Powerless-females-bottom (1258ms; 95% CI[1143, 1373]) vs. powerless-females-top (1289ms; 95% CI[1175, 1404]), $p = .66$. 4) Powerless-females-bottom vs. powerless-males-bottom (1288ms; 95% CI[1186, 1390]), $p = .73$.
Study 4	To experimentally determine whether social status cues affect the strength of spatial simulations by directly priming participants with social-status items on some of the spatial task trials.	1) Powerful-males-top (1010ms; 95% CI[895, 1126]) vs. powerful-males-bottom (1054ms; 95% CI[938, 1170]), $p < .09$. 2) Powerful-males-top vs. powerful-females-top (1079ms; 95% CI[957, 1201]), $p < .06$. 3) Powerless-females-bottom (1048ms; 95% CI[932, 1165]) vs. powerless-females-top (1033ms; 95% CI[916, 1150]), $p = .86$. 4) Powerless-females-bottom vs. powerless-males-bottom (1047ms; 95% CI[934, 1160]), $p = .99$.
Study 5	To replicate the results of Study 1 when participants responded with vertically arranged keys on the keyboard.	1) Powerful-males-top (979ms; 95% CI[896, 1062]) vs. powerful-males-bottom (996ms; 95% CI[912, 1079]), $p = .85$. 2) Powerful-males-top vs. powerful-females-top (1040ms; 95% CI[931, 1150]), $p = .45$. 3) Powerless-females-bottom (1093ms; 95% CI[992, 1195]) vs. powerless-females-top (1112ms; 95% CI[1010, 1214]), $p = .83$. 4) Powerless-females-bottom vs. powerless-males-bottom (1125ms; 95% CI[1033, 1216]), $p = .86$.
Study 6	To replicate the results of Study 1 when participants responded with horizontally arranged keys on the keyboard.	1) Powerful-males-top (1010ms; 95% CI[865, 1155]) vs. powerful-males-bottom (1012ms; 95% CI[866, 1157]), $p = .99$. 2) Powerful-males-top vs. powerful-females-top (1002ms; 95% CI[855, 1149]), $p = .99$. 3) Powerless-females-bottom (1043ms; 95% CI[898, 1188]) vs. powerless-females-top (1037ms; 95% CI[891, 1183]), $p = 1$. 4) Powerless-females-bottom vs. powerless-males-bottom (1093ms; 95% CI[946, 1240]), $p = .31$.

Summary of results

The results partially supported my hypotheses. As predicted, I found spatial simulations of power when power was blended with the social category of gender. Specifically, attributing high power to the male gender involved spatial simulations (i.e., a stereotype-consistent choice), but that was not the case when high power was attributed to the female gender (i.e., stereotype-inconsistent choice). This was demonstrated by significantly faster categorisations of males as powerful than females as powerful at the top. My data further indicated that the above effects involved simulations on the vertical dimension, as males as powerful were chosen faster at the top than at the bottom of the screen. This pattern was not true for females. On the one hand, the findings are consistent with Schubert's (2005) results showing that thinking about power involves spatial processing – *powerful* being simulated at the top as opposed to bottom. On the other hand, this process did not occur when participants reasoned in counter-stereotypic ways. The findings of Studies 1 - 6 suggest that spatial simulations are moderated by stereotypic associations. Overall, I found that such simulations were independent of implicit attitudes towards gender and social status or rationality. Except that spatial simulations of males as powerful at the top as opposed to females as powerful in the same location involved a stronger implicit link between rationality and top vertical positions.

Overall, my initial studies indicated that people spatially simulated specific social categories on the vertical dimension when they thought about them as powerful. However, such simulations are moderated by stereotypic links, as I found that only powerful-males were spatially simulated at the top. In the next chapter, I aimed at replicating these findings with another research method to support the above conclusions.

Chapter 3: Measuring spatial simulations with pupillometry

Overview

In Chapter 3, I present a replication study that tested the robustness of the spatial simulation effect of males at the top when they are judged as powerful. To investigate this, I applied a physiological measure of pupillometry. First, to validate whether the recording of pupil size could be used in investigating spatial simulations of power, in one block, I asked participants to complete the task designed by Schubert (2005). In this task, pairs of powerful and powerless groups were presented on the screen. Participants were instructed to identify as quickly as possible which group was *powerful*. Second, in another block, I conducted a replication of my previous design (Study 1). That is, I presented participants with pairs of gendered names (one at the top, the other at the bottom). The task was to select the *powerful* person as fast as possible. During both blocks, I recorded participants' pupil size change across all trials. Consistent with my previous studies, I found that participants' pupil size was significantly increased when they selected a powerful group at the bottom (an incongruent spatial location) as compared to the top. The same was observed when participants indicated male names as powerful when they appeared at the bottom versus top. Also, irrespective of participants' judgement of a powerful person, increased pupil size was noted on trials when male names were presented at the bottom and female names at the top. As pupil dilation is associated with surprise and higher cognitive load, in my experiment, participants seemed to exhibit more surprise and cognitive load on incongruent (powerful-bottom/powerful-males-bottom) as opposed to congruent trials (powerful-top/powerful-males-top). These finding support my results from the previous studies indicating that spatial simulations occur when people reason about power and gender in stereotype-consistent ways.

My initial experiments investigating spatial simulations in the mental representation of stereotypic associations between gender and power were based on reaction-time methodology. Of course, this methodology is widely used in investigating social cognitive processes, as response times in cognitive tasks are a valid index of cognitive load or effort exerted, ease of processing, task demands or difficulty (see Fazio, 1990, for an overview). My tasks were designed in a way that participants could self-generate their judgements and there were no objectively correct or incorrect questions. The only constraint that participants needed to consider was their own subjective perceptions of power in the context of gender. They could either rely on social stereotypes of males as powerful and females as powerless or contradict them. As to spatial hypotheses about the location of such stereotyped concepts within mental representations, shorter response latencies to certain categories would indicate consistency between participants' representation of presented concepts and the presentation of stimuli on the screen. For example, faster reactions to males as powerful when their names appeared at the top could indicate the ease of processing. Such processing advantage most likely resulted from participant's internal representation of gender stereotypes – males are more powerful than females.

I was motivated to further investigate spatial simulations with another research method to test whether the effects could be reliably captured by a non-behavioural research approach, namely physiological recording of pupil size. Overall, I aimed to test whether another method could complement the reaction-time findings. For this purpose I chose pupillometry. Pupillometry is a study of pupil size change that can be used to investigate implicit cognitive processes when external factors influencing pupil activity are kept constant (i.e., luminance). My use of this technique was motivated by two reasons. First, there is evidence indicating that pupil size increases are associated with expectancy violation. Research suggests that pupillary responses can indicate surprise in response to an unexpected event in a social decision-making task (e.g., Sleegers, Proulx, and van Beest, 2015). Second, pupil size measure can be used to precisely measure how much effort people exerted to process certain stimuli after making a social judgement (van der Wel and van Steenbergen, 2018) Hence, in the current chapter, I present a study that attempted to address these issues by employing pupillometry. Below, I review research studies that examined how pupil size is related to human mental processing.

Pupillometry as a measure of cognitive effort

Research investigating cognitive processes using pupillometry began in the 1960s. Initial studies on pupil dilation were conducted by Hess and Polt (1964), who studied cognitive effort. They found that increasing difficulty of mathematical problems that people needed to solve was associated with increases in their pupil size. Building on those studies, Kahneman and Beatty (1966) provided further evidence indicating that cognitive load is associated with increased pupil dilation. In their experiment, five participants completed a memory task that varied in difficulty. The task involved memorising and immediately recalling a series of digits, nouns, as well as transforming digits (i.e., calculating a sum of digit strings). Increased pupil size was noted as a function of task difficulty and amount of material presented. Presenting more items for an immediate recall was also associated with higher pupil dilation. Increased pupil size was found when participants processed digits for transformation and this effect was bigger than the effect noted in the simple recall tasks. Overall, these studies demonstrate that pupil diameter increases with task difficulty and cognitive load.

Many more recent studies reported in the literature support such conclusions. In a recent review, van der Wel and van Steenbergen (2018) conclude that increases in pupil size are most likely tracking changes in cognitive effort that people exert while performing cognitive tasks. They note that pupil dilation is consistently detected across a number of different cognitive tasks and their level of difficulty, with easier tasks being associated with smaller pupil size. These findings were based on studies using tasks involving working memory operations, attentional control, or inhibition of distractors: arithmetic problems (Ahern & Beatty, 1979), the Sternberg task concerning memory (van Gerven, Paas, van Merrienboer, & Schmidt, 2004), or go/no-go task (e.g., Reinhard & Lachnit, 2002). Further, similar findings were obtained in the context of conflict paradigms involving incongruent trials, where the target needs to be detected by inhibiting the distractor stimulus, and congruent trials, where no such inhibition is required (e.g., Stroop, flanker, and Simon tasks). The studies using these tasks indicated that pupil dilation is significantly increased when people complete *incongruent* trials as opposed to *congruent* ones. (e.g., Laeng, Orbo, Holmlund, & Miozzo, 2011; Cohen, Moyal, & Henik, 2015). Overall, it appears that pupillometry is a valid measure of cognitive load.

In terms of physiological underpinnings, changes in pupil size due to cognitive effort are associated with the activity of the two components of the autonomic nervous system that influence the functioning of the iris, which controls the pupil diameter (Beatty & Lucero-Wagoner, 2000; Steinhauer, Siegle, Condray, & Pless, 2004). The iris is controlled by the sphincter muscle, which is affected by parasympathetic activity, and the dilator muscle, which is associated with the sympathetic activity. Increased inhibition of parasympathetic activity and enhanced sympathetic activity directly cause the pupil to dilate. Specifically, enhanced sympathetic activity is directly related to pupil dilation by influencing the dilator muscle. In turn, increased sympathetic activity is associated with attentional vigilance, alertness to salient stimuli, attentional focus, as well as cognitive effort (Fairclough & Mulder, 2011).

Such mechanisms are also correlated with brain activity. A state of increased attentional arousal is linked with neural brain activity of the locus-coeruleus norepinephrine (LC-NE) system, which correlates with pupil dilation (see Aston-Jones & Cohen, 2005, for an overview). The LC-NE system activates in response to presentation of salient stimuli, to facilitate their processing. This happens because the task-evoked phasic activity of LC produces a rapid release of NE to the parietal and frontal areas involved in cognitive control, for example, the anterior cingulate which is involved in processing task demands (Murphy, Robertson, Balsters, & O'Connel, 2011). Such release increases neuronal gain in the relevant brain areas, aiding stimulus processing (Aston-Jones & Cohen, 2005).

Pupillometry in expectancy violation

The reviewed studies suggest that pupil diameter is associated with task demands or cognitive effort needed to perform a cognitive task. In relation to such findings, another line of research suggests that pupil dilation can reflect expectancy violation or surprise. This is because unexpected events are related to heightened physiological arousal, the fight-or-flight response, and so elicit increased vigilance and attentional focus. As one example, Sleegers et al. (2015) demonstrated that expectancy violation associated with the presentation of anomalous playing cards (e.g., king of hearts presented in black instead of red) evoked larger pupil size than the presentation of normal cards. Moreover, this research also demonstrated that such effect was moderated by participants' extremism in beliefs about societal issues, such that endorsing more extreme values involved reduced pupil dilation. This is because extremism diminishes conflict arousal. Not only do such findings support the previous literature in indicating that pupil size is associated with expectancy violation, but also they indicate that social extremism might modulate physiological arousal. Such arousal can be reliably measured with pupillometry. This was also supported in a study where participants were asked to complete a gambling task and predict which playing card would be higher in order to receive a reward (Preuschoff, Hart, & Einhauser, 2011). It was found that uncertainty or expectations of reward themselves were not predicted by pupil size. Instead, increased pupil size was noted when participants committed an error in judging which cards were higher, signalling surprise (pupil dilation correlations with surprise were also demonstrated by Kloosterman et al., 2015 or Proulx et al., 2017).

Present study: Pupil size in stereotyping

Overall, the previous literature indicates that presentation of unexpected events and cognitive effort can be reflected in increased pupil diameter. There is also evidence that pupillary activity can be modulated by extremity of social beliefs suggesting that pupil size can be used to investigate social psychological processes. Also, subtle cues that violated people's expectations (e.g., anomalous playing cards), which were processed unconsciously, were sufficient to elicit increased pupil size. If such subtle cues impact pupillary activity, it is possible that presenting concepts in their metaphorically "incongruent" spatial positions (powerful person presented at the bottom as opposed to the top) would also be associated with initial expectancy violation responses and hence increased pupil size. Subsequently, selecting an inconsistent stimulus as powerful (e.g., female names) should be associated with increased cognitive effort (Aston-Jones & Cohen, 2005).

Therefore, in the present study, I investigated whether pupillometry can be used as a proxy of cognitive conflict and response-locked cognitive effort in research on sensorimotor processes involved in abstract thinking. To do that, I designed an eye-tracking experiment involving two stages. In the first stage, I presented participants with Schubert's task (2005, Study 2) in order to test whether pupil size changes can indeed parallel the effects obtained using the reaction-time methodology applied by Schubert. However, because Studies 1- 6 did not detect significant spatial effects in the powerless condition, I limited the task to selecting the powerful member of the pair. Participants were presented with manipulated vertical positions (top versus bottom) of generic powerful (*master*) and powerless groups (*servant*) and asked to detect as quickly as possible the powerful group on the screen.

Given that mental representations of power involve spatial correlates, such that powerful individuals are simulated at the top and powerless at the bottom (Schubert, 2005) and cognitive conflict and effort are indexed by pupil dilation (Aston-Jones & Cohen, 2005; Sleegers et al., 2015; van der Wel and van Steenbergen, 2018), I hypothesised that presenting participants with powerful groups at the bottom of the screen versus the top would be associated with an initial increase in pupil size due to surprise before the response would be made. This would be observed in the *pre-response* period of a trial, that is, before participants chose the powerful person. Likewise, having responded to an incongruent trial (powerless-bottom) in contrast to a congruent trial (powerful-top) would be associated with increased response-locked cognitive effort leading to continuous pupil dilation. This would be observed in the *post-response* period.

In the second stage of the experiment, I presented participants with the spatial task adapted from Study 1 (Chapter 2). Again, I asked participants to select as quickly as possible the powerful person from a pair. As in Study 1, in the present experiment, participants were presented with pairs of gendered names. Again, I manipulated the trial type. On males-top trials, participants were presented with male names presented at the top and female names at the bottom and vice versa on females-top trials (see Figure 5). Participants' categorisations of male and female names as powerful in each vertical position were recorded and constituted a quasi-experimental variable (see Chapter 2). On the basis of my previous studies presented in Chapter 2, I predicted that participants would exhibit increased initial arousal associated with cognitive conflict, and so increased pupil dilation associated with the presentation of female-top trials as opposed to males-top trials in the *pre-response* period. They would then demonstrate further pupillary dilation associated with the response, only *after* indicating females as powerful at the top as opposed to males as powerful in the same location, i.e., in the *post-response* period. I speculate that such higher cognitive effort, as reflected by increased pupil dilation, should be observed in response to potential inhibition of the automatically activated existing association between powerful-males. Specifically, pupil dilation would indicate an attempt to control the spatial simulation involved in the mental representation of gender stereotypes. I also predicted that the presentation and selection of male names when they appeared at bottom versus top would be associated with increased pupil dilation. In my previous studies, I did not find differences in simulating females as powerful at the top versus the bottom, hence, I predicted no differences in cognitive conflict or effort in those cases.

Both stages of the experiment, that is, the Schubert and gender blocks, were presented to all participants. Their order was counterbalanced across participants. Within each block, I measured participants' pupil size. To acquire enough pupil size data in the post-response period, I presented each trial for 4000ms, however, I encouraged participants to respond as quickly as they could. In the gender block, I also recorded how often participants picked each gender in each vertical position. In both blocks, each trial lasted 4000ms. Therefore, if participants did not respond within that period, their answers were not recorded (such responses constituted less than 1% of the data).

Method

Participants

Sixteen Cardiff University undergraduate students participated in the experiment (13 females, one non-binary, mean age = 19.44; the approximate sample size for medium effect size was derived from Laeng et al., 2011). They received course credit for their participation.

Materials

For the gender block trials, I adapted the spatial task materials from my previous experiment (see materials of Study 1). That is, I presented five pairs of gendered names as stimuli, such that on half of the trials male names were positioned at the top and female names at the bottom or vice-versa on the other half of the trials. In turn, in the Schubert block, I presented five pairs of powerful and powerless groups – I randomly selected five powerful and five powerless groups from the list of items used by Schubert (2005, Study 2). To standardise the conditions of both gender and Schubert blocks, I also randomly paired each powerful group with a powerless group and presented the same pairs across the experiment (master-servant, coach-athlete, employer-employee, officer-soldier, and boss-secretary). The vertical position of the powerful-powerless pairs was also manipulated in the same ways as in the gender block. Each block included 40 trials, therefore, there were 80 trials in total.

To measure participants' eye movements and pupillary activity, I used a Tobii X3-120 screen-based and non-invasive eye tracker that samples data at the speed of 120Hz (Tobii, Stockholm, Sweden). The eye tracker was integrated in a 17'' screen with a resolution of 1920 x 1080 pixels. The pupil size output is corrected by an in-built algorithm that compensates for eye movements and changes in the angle at which pupil size is recorded.

Design

I used a within-participants design, so that each participant contributed to both the gender and Schubert blocks. I asked participants to select the powerful person on the screen. The order of blocks was counterbalanced between-participants and trials were randomly presented within each block. The trial type in the gender block (males-top/females-top) and the vertical position of powerful and powerless groups (powerful-top/powerless-bottom) were manipulated within-participants. In the Schubert block, I measured participants' pupil size to the trials in which they made a correct response (i.e., they indicated a powerful group as powerful). In the gender block, I also recorded pupil size on all types of trials (male/female names categorised as powerful at the top or bottom).

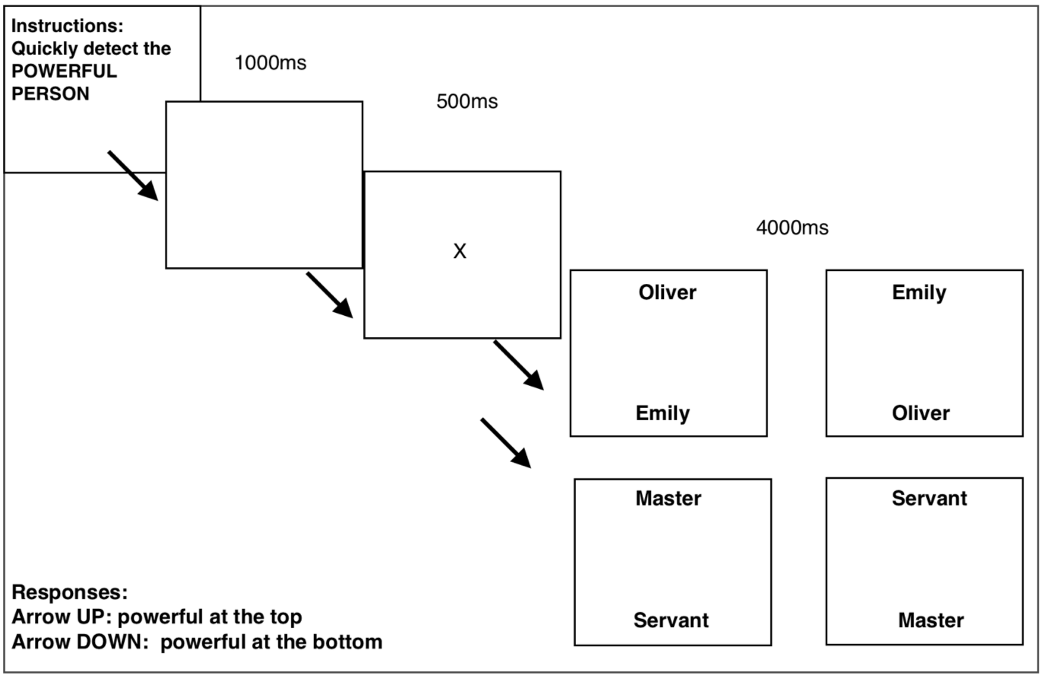
I used E-Prime 2.0 software to present all stimuli, record responses, and measure pupillary activity registered by the eye tracker (Psychology Software Tools, Pittsburgh, PA).

Procedure

After signing the consent form, participants were instructed sit in front of the computer screen (approximately 67cm away) and were informed that first they would take part in an eye-tracking calibration procedure. During this procedure, an eye tracker collects information about properties of participant's eyes (e.g., light refraction, cornea, fovea position) in order to create an internal 3D eye-model that enables accurate measurement of the participant's eye gaze and pupil size in the main experimental task (see Tobii Group, 2015). In this procedure, participants are presented with a red dot against a grey background and instructed to follow the dot with their gaze. The dot moves smoothly from one point to another and is presented at several different locations on the screen (e.g., in the middle of the screen/middle right/left or top and bottom right/left). If the calibration is unsuccessful (e.g., due to a failure to follow the dot), the procedure needs to be repeated. The quality of calibration is calculated by E-Prime software and presented at the end of the calibration procedure.

Subsequently, participants were instructed to complete the experimental task and quickly categorise the powerful person on the screen for both the Schubert and gender trials (see Figure 5). The instructions were the same as for the previous studies reported in this thesis, that is, they included a definition of a socially powerful person. Participants were tested in a dark room. The experiment lasted 45 minutes.

Figure 5. Spatial task procedure.



Results

Pupil size pre-processing. First, all the gaze data that were missing due to blinks or incorrect recordings (as computed and indicated by E-Prime software) were coded as missing values. These constituted 22% of the data in the Schubert block and 18% in the gender block. Second, I combined the pupil data from the left and right eye into one pupil size score that represented the mean pupil size for both eyes. Following the procedure used by Proulx et al. (2017), I subsequently filtered the pupil size signal with a repeated median regression filter[14] (robfilter package by Fried, Schettlinger, & Borowski, 2014) in statistics software R (R Core Team, 2016) to smooth implausible pupil recordings. Such invalid measurements might be due to participants' head movements during the task, as no chin rest was used. First, the filter determines the slope/trend of pupil size increases or decreases over time within a specified inner window width (determined by the samples taken to estimate the slope). Once the slope is identified and estimated, the median is calculated for the slope estimations for the whole specified outer window. The inner and outer widths should be specified on the basis of the visual inspection of the data for each participant and each trial. The most appropriate filter for my data required the inner width of 15 and outer width of 25 samples.

Once the pupil size signal was smoothed, the missing signal that was not likely associated with eye blinks (as its duration was too short, i.e., fewer than four data entries, which equal to 33ms) was corrected by a linear interpolation.[15]

Subsequently, I removed artifacts associated with blinks. Blinks are recorded as missing data by the eye tracker, however, they also lead to incorrect measurements around 100ms before they occur (as participants' eyelids close, the eye tracker fails to record the entire pupil). Also, incorrect measurements are likely to be recorded 200ms post-blink. This is caused by opening of eyelids as well as an initial constriction of the pupil that occurs immediately after each blink (Lin et al., 2018). Therefore all pupil recordings 100ms pre-blink and 200ms post-blink were removed. Finally, the missing

[14] In general, median filters inspect time-series signal entry by entry within a pre-determined window (a window indicates a pattern of adjacent values) and replace the data points within that window with a median of entries adjacent to the window.
[15] The linear interpolation replaces missing signal with the mean of neighbouring valid signal measurements.

signal due to blinks was filled in by a linear interpolation. The missing signal recorded for more than 500ms was not interpolated, as blinks do not last longer than this period. Overall, after pre-processing, the estimated data loss was reduced to 11% in the Schubert block and 9% in the gender block.

Finally, to correct for baseline differences in pupil size between participants, I calculated the mean pupil diameter during 100ms of the fixation period that occurred just before the target trial period presentation following Mathot et al.'s (2018) suggestions (the target trial period refers to the trial event when pairs of powerful-powerless groups and female-male names appeared on the screen). This was done for each participant and each trial. Then, I subtracted that mean score from each subsequent pupil size measurement during the target trial period.[16] After the baseline corrections, all pupil size values that were closer to 0 represented a normal pupil size for participants before each trial began.

Schubert block. General analysis. To analyse participants' pupil diameter across trials where they correctly detected powerful groups, I first removed incorrect or missing responses (participants were not allowed to respond after 4000ms). The incorrect responses constituted the trials when participants indicated a powerless group (both at the top or bottom) as a powerful one. The trials with missing or invalid responses accounted for 3% of all responses.[17] Subsequently, as I collected 120 samples of the pupil size data per second (480 samples in each trial), I computed the mean pupil size for each participant, trial, and trial type (powerful groups shown at the

[16] As the literature does not report one valid way of pupil pre-processing techniques (e.g., Cavanagh, Wiecki, Kochar, & Frank, 2014; Lin et al., 2018; Mathot, Fabius, van Heusden, & van der Stigchel, 2018; Proulx et al., 2017), I combined different smoothing and filtering techniques described in the literature that best fit my data in terms of signal quality, signal loss, eye-tracker speed, and trial length. I reported only one technique, however, using different window widths in median filters or not applying blinks corrections led to similar trends in the data analyses.
[17] As the error rate accounted only for 3% of trials, I did not analyse pupil size associated with them.

top and powerless at the bottom or vice versa). I then calculated the grand mean of reaction time (*M* = 1379, *SD* = 429) across all participants and trials. I used the mean response to determine a general pre-response and post-response period across trials.[18] Finally, to visually inspect the pattern of the mean pupil size change over the trial period, I calculated mean pupil size for time bins every 250ms (see Figure 6).

Once the above steps were achieved, I estimated a linear mixed model (see Appendix 1, Study 7, for the method of modelling the effects) to test whether there was a difference in pupil size in the powerful-top and powerful-bottom trials. First, I introduced the response period (pre-response and post-response) and position (powerful groups at the top or bottom) as factors. Second, I analysed the whole trial period (500-4000ms). I excluded the first 500ms after the trial onset, as this period was associated with an initial constriction of pupil due to the light reflex response, that is, a presentation of a stimulus on the screen resulting in a change of luminance (Prehn et al., 2008).

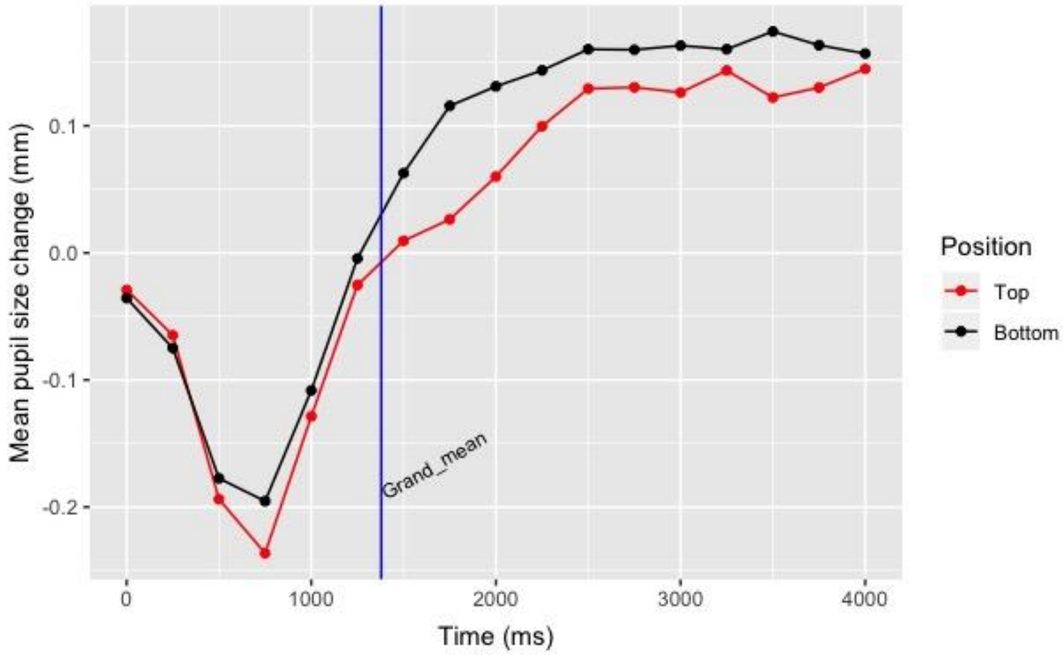
As expected, I found that the fixed effect of position was significant across the whole trial period (500-4000ms), $F(1, 1125) = 4.00$, $p < .05$, $dz = .12$,[19] indicating that the pupil dilation was higher in the case of the powerful-bottom trials, (*M* = .001mm; *SE* = .03, 95% CI[-.07, .07]), as opposed to the powerful-top trials, (*M* = -.030mm; *SE* = .03, 95% CI[-.10, .04]), suggesting more arousal in the powerful-bottom trials. The fixed effect of the response period was significant, $F(1, 16) = 42.20$, $p < .001$, $dz = .87$, with significantly larger pupil size in the post-response period (*M* = .110mm; *SE* = .04, 95% CI[-.04, .18]), as opposed to the pre-response period, (*M* = -.14mm; *SE* = .03, 95%

[18] Before splitting the data according to the response period, I trimmed RT data according to the Tukey criterion. No range restriction was applied because participants were allowed to respond within the period of 4000ms. This was done for both the Schubert and gender block.
[19] The p-value for the main effect of position varied from $p < .001$ to $p < .05$ depending on different filter parameters or when no blink correction was used. It is possible that different pre-processing corrections, which led to different percentages of data loss, resulted in more or less underpowered results.

CI[-.24, .05]). This effect indicates that the pupil was larger after participants made their responses. However, the interaction between the response period and position was not significant, $F(1, 1125) = .15$, $p = .70$, suggesting that the pattern of differences in pupil size across powerful-top/powerful-bottom trials was stable across the whole trial period.

Schubert block: Time-stamp analyses. As can be seen in Figure 6, pupil size differences between trial types (powerful groups at the top and bottom) were the most pronounced in the case of time stamps at 750ms in the pre-response period and 1500ms, 1750, and 2000ms in the post-response period. Hence, I conducted additional analyses of pupil size differences at those time bins. First, at 750ms, I found that participants' pupil size was marginally more dilated in the powerful-bottom trials (M = -.20mm, SE = -.05, 95% CI[-.31, -.09]) than powerful-top trials (M = -.24mm, SE = -.05, 95% CI[-.34, -.13), $F(1, 556) = 3.35$, $p = .07$, $dz = .58$. In the post-response period, at time-stamp of 1500ms, participants' pupil diameter was significantly increased in the powerful-bottom (M = .05mm, SE = .04, 95% CI[-.02, .15]) than powerful-top trials, (M = .01mm, SE = .04, 95% CI[-.08, .09]), $F(1, 501) = 6.27$, $p < .01$, dz =.28. The same pattern and the biggest difference in pupil size between powerful-top, (M = .02mm, SE = .04, 95% CI[-.06, .11]), and powerful-bottom trials, (M = .11mm, SE = .04, 95% CI[.03, .20]), was observed at 1750ms time-stamp, $F(1,499) = 13.80$, $p < .001$, $dz = .57$. The same was true for 2000ms time-stamp, $F(1,499) = 12.36$, $p < .001$, $dz = .50$ (for powerful-top: M = .05mm, SE = .04, 95% CI[-.03, .13], and powerful-bottom: M = .13mm, SE = .04, 95% CI[.05, .22]). The significant differences between powerful-top and -bottom trials observed in the post-response period indicate that cognitive effort was higher on the powerful-bottom trials than -top trials after correctly categorising powerful groups. It is possible that post-response cognitions are associated with reconciling one's own response in the context of the inconsistent presentation. Finally, the differences in pupil size at other time stamps in the pre-response and post-response periods were not significant, *ps* > .14.

Figure 6. Mean pupil size change as a function of presenting powerful groups at the top (powerless at the bottom) or vice versa over the trial period (4000ms). The vertical line indicates the mean response time across all participants and conditions (*M* = 1379ms, *SD* = 429) in the Schubert block. Participants were instructed to look for the powerful person.



Gender Block. Next, I computed the mean pupil size for each participant, trial, trial type (males-top versus females-top), and gender that was indicated by participants to be powerful (gender choice: male versus female names). Then, I calculated the mean response time (*M* = 1415, *SD* = 666) across all participants and trials to obtain a reference in order to analyse participants' pupil size in the pre-response and post-response periods.

Gender block. Pre-response period. First, I tested whether just presentation of females-top trials versus males-top trials (regardless of participants' gender choice later) would involve increased pupil size in the context of the task instructions (i.e., categorising the powerful person). If the pupil size increases due to unexpected stimuli presentation, such increased dilation should be present in the pre-response period regardless of participants' choice. However, if the whole pre-response period was associated with increased pupil dilation, then it is most likely that such differences are simply related to cognitive load associated with processing of information. Figure 7 suggests that the biggest difference in pupil size was observed in the case of 750ms time stamp. As hypothesised, participants' pupil dilation was significantly higher when they were presented with females-top trials (M = -.14mm, SE = .05, 95% CI[-.24, -.04]), as opposed to males-top trials, (M = -.19mm, SE = .04, 95% CI[-.28, -.08]), $F(1,590)$ = 4.81, $p < .03$, dz = .43, signalling more surprise on females-top trials. Also, as expected, this effect was independent of participants' later response within the trial, which is demonstrated by the lack of significant interaction between gender choice and trial type, $F(1,593) = 1.76$, $p = .19$. As demonstrated by Figure 7, and confirmed by additional analyses, the differences in pupil sizes across the trial types within the whole pre-response period (500-1415ms) as well as individual time-stamps of 500ms and 1000ms were not significant, $ps > .10$. Such findings indicate that there was initial increase in pupil size at 750ms in the females-top trials, suggesting that presenting females at the top was likely to evoke cognitive conflict or surprise. This is also supported by the results showing that such increase in pupil size was not associated with the response that participants made later within the trial. After 750ms, pupil size within both conditions increased to the same extent most likely as a function of general cognitive load associated with decision-making.

Furthermore, as indicated by Figure 8a, it seems that participants demonstrated increased pupil size when they were presented with males-top trials and subsequently indicated females as powerful. The differences were most pronounced at 500ms and 750ms and are still visible in the post-response period. The same difference, however, is not demonstrated by Figure 8b, suggesting no differences in pupil size between choices of females (at the top) or males (at the bottom) on females-top trials. To test whether these differences were meaningful in the context of subsequent decision-making, I conducted additional analyses. Post-hoc comparisons indicated that the difference in pupil size between choices of males when they appeared at the top and females when they appeared at the bottom at 500ms was marginally significant, $t(596) = -1.80$, $p = .07$, $dz = .98$, such that pupil dilation at 500ms was increased when females were subsequently chosen at the bottom ($M = -.12$mm, $SE = .05$, 95% CI[-.28, -.06]), in contrast to choosing males at the top ($M = -.25$mm, $SE = .05$, 95% CI[-.32, -.11]). This is consistent with the previous literature suggesting that when people decide to control their initial tendencies, like in my task categorising females as powerful at the bottom (when a choice of males or top is available), they need to inhibit the powerful-top association and/or the stereotype associated with males as powerful first (van der Wel & van Steenbergen, 2018). This in turn involves higher pupil size. The same, however, was not true at 750ms, as the interaction between gender choice and trial type at that time-stamp was not significant, $F(1, 590) = 1.76$, $p = .19$. The same differences in pupil size later on within the trial were not significant at individual time stamps, $ps > .19$. Finally, comparisons between choices of males at the bottom versus females at the top on females-top trials were not significant, $ps > .33$.

Figure 7. Mean pupil size change as a function of presenting males-top and females-top trials in the pre-response period of the gender block. The vertical line indicates the mean response time across all participants and conditions (*M* = 1415ms, *SD* = 666). Participants were instructed to look for the powerful person.

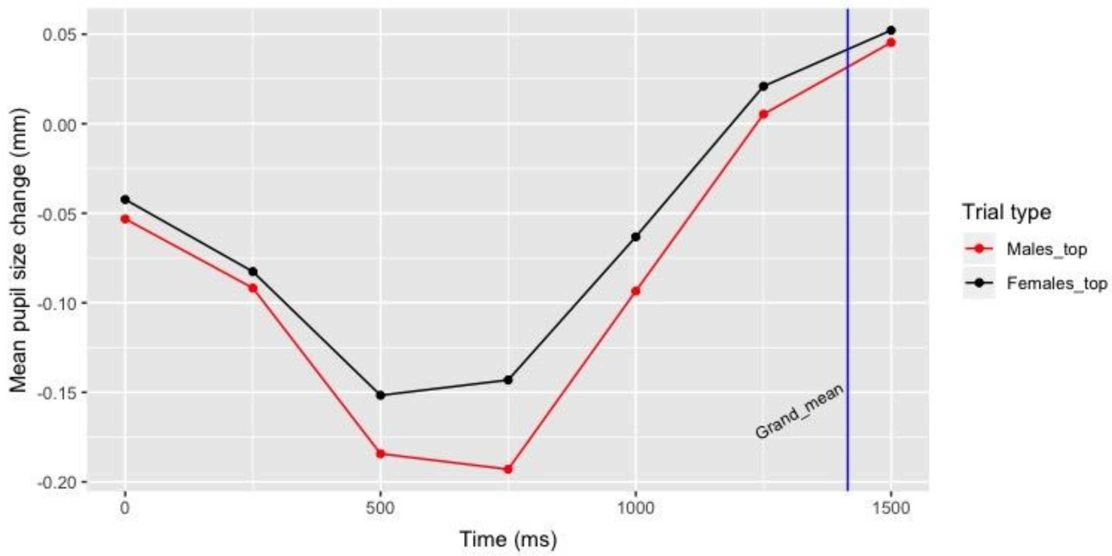
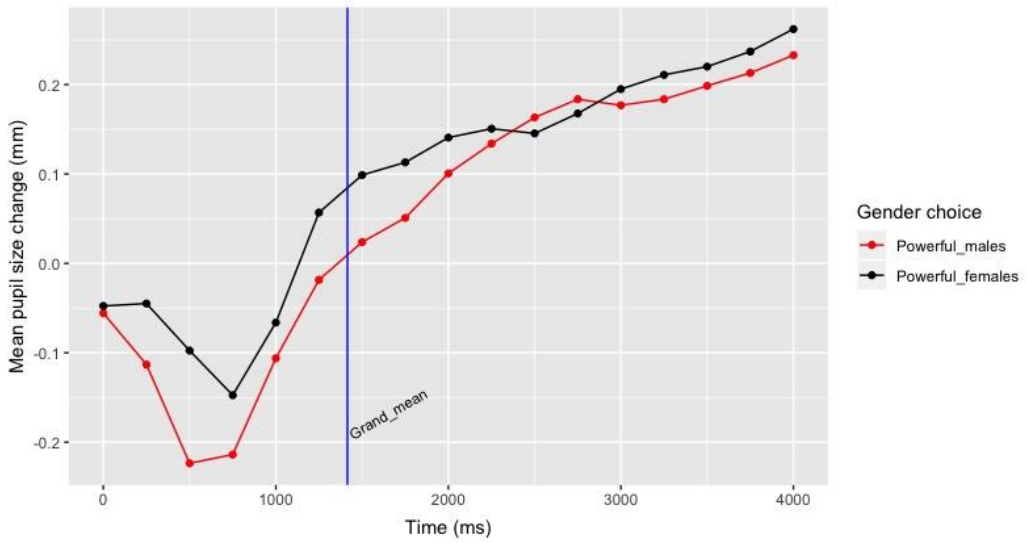
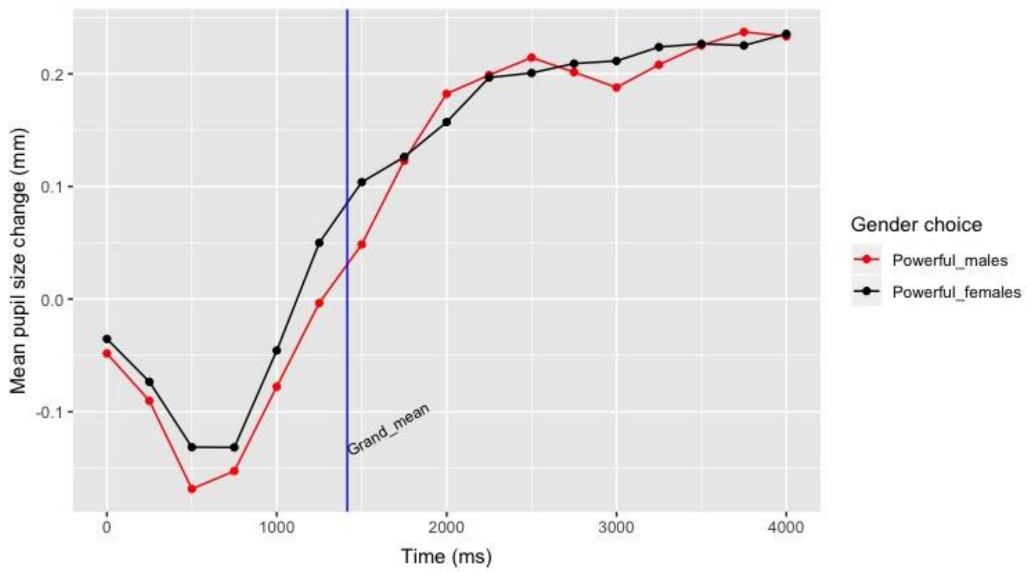


Figure 8. Mean pupil size change as a function of presenting (a) males-top trials and choices of male and female names on that trials (i.e., males at the top; females at the bottom; both chosen as powerful) and the same choices are presented for (b) females-top trials over 4000ms. The vertical line indicates the mean response time across all participants and conditions (*M* = 1415ms, *SD* = 666).

(a) Males-top trials



(b) Females-top trials



Gender block. Entire-trial period analysis. Subsequently, I analysed participants' pupil dilation across the entire trial period (500-4000ms). I found that the main effects of gender choice, *F*(1,13) = .76, *p* = .40, *dz* = .07, and trial type, *F*(1,1168) = .97, *p* = .32, *dz* = .07, were not significant. As in the Schubert block, the main effect of response period was significant, *F*(1,15) = 37.01, *p* < .001, *dz* =.86, with increased pupil dilation in the post-response period, (*M* = .17mm, *SE* = .04, 95% CI[.07, .26]), than in the pre-response period, (*M* = -.09mm, *SE* = .04, 95% CI[-.17, -.01]). As expected, the two-way interaction between trial type and gender choice was significant, *F*(1,15) = 37.01, *p* < .001. On trials where participants selected males as powerful, their pupil size was significantly increased when the male name appeared at the bottom (*M* = .05mm, *SE* = .03, 95% CI[-.01, .13]) as compared to the top, (*M* = .01mm, *SE* = .03, 95% CI[-.06, .07]), *F*(1,1190) = 3.15, *p* < .01, *dz* = .19 (see Figure 9). However, the same was not true on trials where participants indicated females as powerful when they appeared at the bottom of the screen (*M* = .05mm, *SE* = .03, 95% CI[-.03, .17]) compared to top (*M* = .06mm, *SE* = .05, 95% CI[-.06, .14]), *F*(1,1199) = 1.20, *p* = .23, *dz* = .10 (see Figure 10).

Finally, two-way interactions (trial type by response period; and gender choice by response period) and the three-way interaction (gender choice by trial type by response period) were not significant *ps* > .61.

Further, as indicated by Figure 11, participants' pupil dilation was increased when participants indicated females as powerful on females-top trials (*M* = .060mm, *SE* = .05, 95% CI[-.03, .17]) as opposed to males as powerful positions on males-top trials, (*M* = .006mm, *SE* = .04, 95% CI[-.06, .13]), *t*(1205) = -1.96, *p* = .05, *dz* = .18. Consistent with Figure 12, there was no significant difference between pupil dilation when participants indicated bottom positions on males-top trials (i.e., when females as powerful were chosen at the bottom; *M* = .05mm, *SE* = .05, 95% CI[-.01, .19]), and females-top trials (i.e., when males as powerful were chosen at the bottom; *M* = .05mm, *SE* = .05, 95% CI[-.03, .17]), *t*(1206) = -.17, *p* = .86, *dz* =.01. Overall, the above results imply that across the whole trial period, participants exhibited increased pupil size, and hence higher cognitive effort, when they selected a counter-stereotypic stimulus (i.e., females as powerful) at the top as opposed to a stereotypic stimulus (i.e., males as powerful) at the top. Further, there were no differences in pupil size when selecting both counter-stereotypic stimuli and stereotypic stimuli at the bottom suggesting that cognitive load associated with those choices was comparable.

Gender block. Post-response period. Finally, as in the gender block participants were allowed to make any type of choice (males and females as powerful at the top or bottom), I tested whether their responses were associated with higher cognitive effort. Therefore, I conducted the same analysis as presented above, however, this time I analysed participants' pupil size only after they made their response (i.e., in the period > 1415ms). Similar to the analysis of the whole trial period, I found that the main effects of gender choice and trial type were not significant, *ps* > .72. However, as expected, the gender choice by trial type interaction was significant, $F(1, 595) = 5.23$, $p < .02$. The post-hoc comparisons replicated the results of the previous analysis, indicating increased pupil dilation when males were chosen as powerful when they appeared at the bottom of the screen (M = .18mm, SE = .04, 95% CI[.11, .27]) as compared to the top (M = .14mm, SE = .04, 95% CI[.05, .21]), $F(1, 600) = -2.31$, $p < .02$, dz =.16. As expected, the same was not true for females chosen when they appeared at the bottom of the screen (M = .16mm, SE = .05, 95% CI[.08, .31]) versus the top, (M = .18mm, SE = .05, 95% CI[.05, .27]), $F(1, 600) = 1.12$, $p = .26$, dz =.07. Further analyses did not replicate the findings from the entire-period results, indicating that participants' pupil size was not significantly different when participants indicated top or bottom stimuli across males-top and females-top trials, *ps* > .25.

The reported results within both blocks were not associated with participants' gender or counterbalancing of blocks, *ps* > .20.

Figure 9. Mean pupil size change as a function of choosing males as powerful when their names appeared the top and bottom over the trial period (4000ms). The vertical line indicates the mean response time across all participants and conditions (*M* = 1415ms, *SD* = 666). Participants were instructed to look for the powerful person.

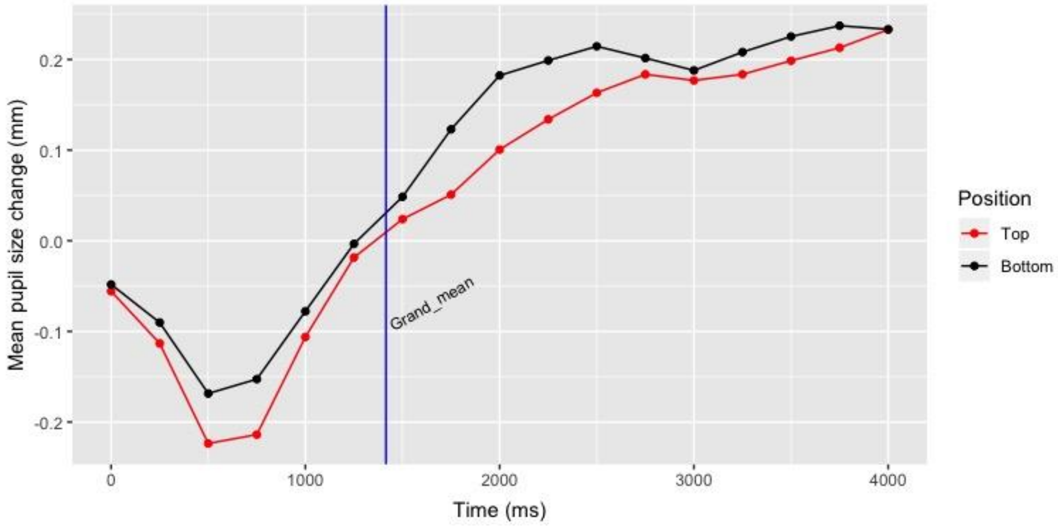


Figure 10. Mean pupil size change as a function of choosing females as powerful when their names appeared at the top and bottom over the trial period (4000ms). The vertical line indicates the grand mean of response time.

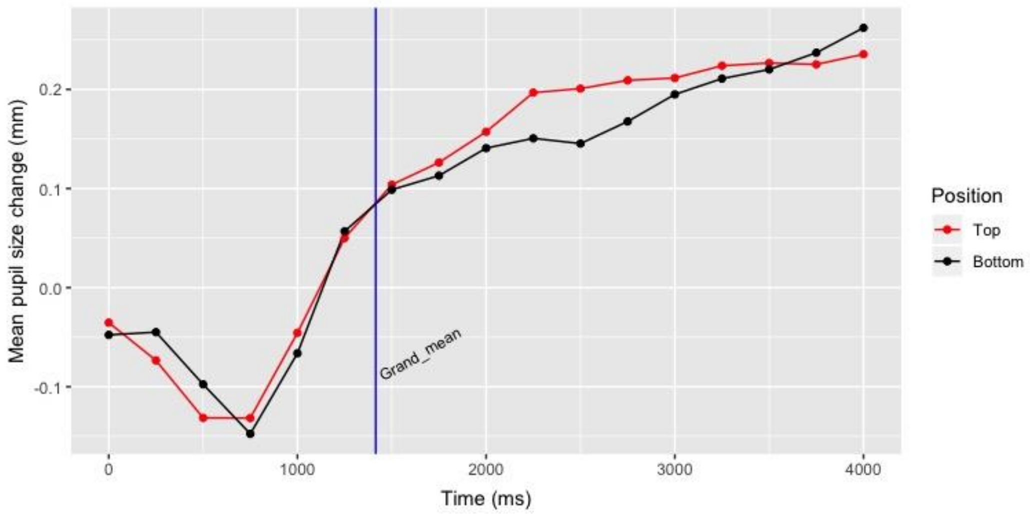


Figure 11. Mean pupil size change as a function of choices of males and females as powerful at the **top** (*Males_top* indicates males-top trial and the choice of males, whilst *Females_top* indicates females-top trials and the choice of females).

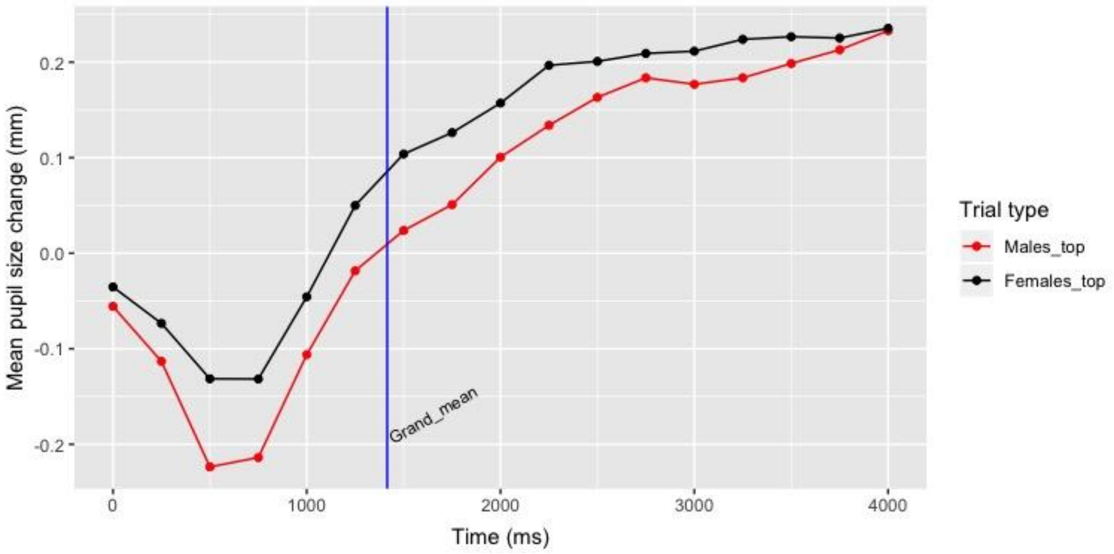
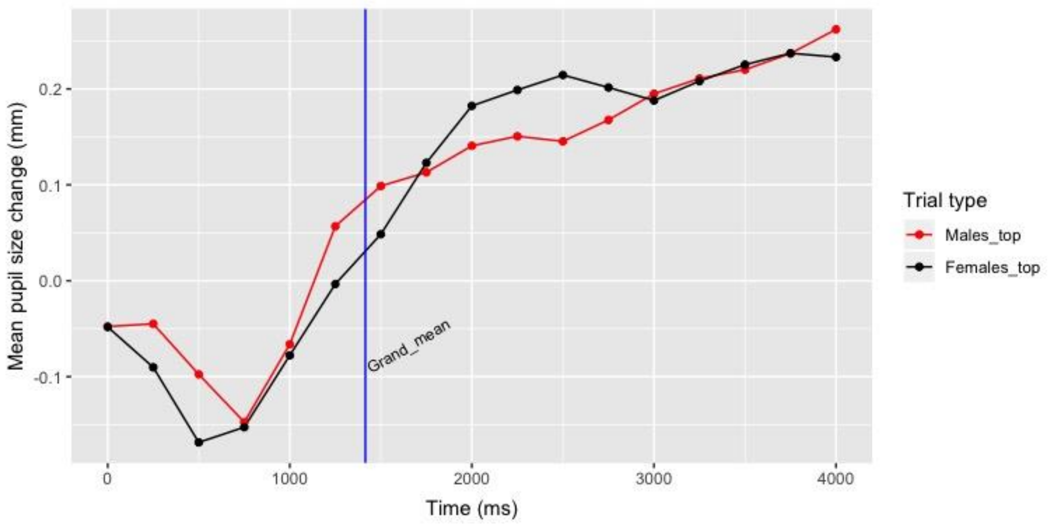


Figure 12. Mean pupil size change as a function choices of males and females as powerful when their names were selected at the **bottom** (*Males_top* indicates choices of females at the bottom, whilst *Females_top* indicates males chosen at the bottom).



Summary of results

To summarise, it appears that pupillometry can be applied to the study of mental representation of stereotypes in the context of sensorimotor grounding as a complementary method to reaction-time methodology. This is because pupillometry can indicate cognitive conflict and mental effort exerted while making judgements. My findings indicate that there is an association between power and verticality, such that powerful groups are associated with top vertical positions, replicating Schubert's results. These findings apply in turn to specific social categories, given that those categories are associated with power via social beliefs. Such findings were based on the measurement of participants' pupil size while they were completing spatial tasks. In those tasks, increased pupillary dilation was associated with the presentation of powerful groups at the bottom. The same pattern was found when participants were presented with female names at the top of the screen compared to male names at the top, when they were asked to categorise the powerful person, suggesting increased cognitive conflict or surprise on females-top trials. At the same time, I observed significantly increased pupil size on trials when participants categorised males as powerful at the bottom versus the top. Such increased pupil size was most likely associated with response-locked cognitive effort. It is important to note that the initial increased cognitive conflict that occurred in the pre-response period (due to the perception of the incongruent mappings of targets [e.g., males-bottom]) might have been associated with the increased response-locked effort that occurred in the post-response period. This possibility should be further investigated in future research. Examining how cognitive conflict and effort interact might provide us with a better understanding of the processes underlying spatial congruency effects in the context of pupillometry.

In my previous studies (1 - 6), only males as powerful were simulated at the top (also in contrast to females) as opposed to the bottom, suggesting that a stereotypic link between males and power involves spatial simulations, hence, more pronounced or concrete thinking. Overall, the response latency findings, indicating faster reactions to males at the top versus bottom only when they are categorised as powerful, converge with the current study, that demonstrated increased cognitive effort on trials when participants selected males as powerful when their names appeared at the bottom of the screen versus the top, as indexed by increased pupil size. Interestingly, I found that counter-stereotypic choices might have been based on the initial experience of unconscious conflict demonstrated at the beginning of the trials when females were presented at the top. Furthermore, only the earliest experience of aversive arousal (at 500ms, but not at 750ms) on females-top trials interacted with counter-stereotypic categorisations later on within the trial.

Altogether, it appears that gender is indeed spatially simulated on the vertical dimension via associations with power. As these results were based on reaction-time methodology and also pupillometry, it appears that such findings are robust. In the subsequent chapter, I aimed at investigating specific mechanisms driving such simulations.

Chapter 4: Dependency of spatial simulations on task features

Overview

In the following chapter, I present three studies that investigated mechanisms of spatial simulations detected in my previous experiments. Specifically, I was interested to examine the dependency of spatial simulations on verticality cues (top and bottom positions) as well as salience of the power context. To examine this, in Studies 8 and 9, I aimed at creating conditions whereby the independent effect of spatial cues on spatial simulation of power could be investigated. Therefore, across two studies, I assigned participants to either select a powerful or powerless person between pairs of names that had the same gender. That is, participants were asked to select the powerful/powerless person between two female names or male names that were vertically displayed (one at the top, the other at the bottom). To summarise, I manipulated gender of the names (female block versus male block) and task condition (powerful versus powerless). I measured how often and how fast participants selected top and bottom positions of the presented targets. In Study 10, I eliminated the power-context by asking participants to determine whether vertically presented (top versus bottom) gendered names (male versus female shown one at a time) were male or female. This was done to test if gender per se was associated with vertical positions. In Studies 8 and 9, I found that participants were equally fast at detecting male and female targets at the top versus the bottom in the powerful condition. The same was true in the powerless condition. In Study 10, participants were equally fast to categorise male and male names at the top and bottom. These results imply that spatial simulations occur only in meaningful and power-relevant conditions.

In Chapters 2 and 3, I investigated the primary research question of the present thesis. Specifically, in seven studies I explored whether abstract thinking about a specific social category would be associated with concrete spatial processing. I provided evidence for the involvement of spatial simulations in abstract thoughts about power and gender in the context of stereotypic associations. The findings from six initial behavioural experiments, as well as physiological results obtained in Study 7 using pupillometry, clearly suggest that the mental representation of males as powerful uses spatial processing, whereby males as powerful are associated with the top vertical position via stereotype fit (males-power). Males were categorised faster as powerful when they were presented at the top of the screen as opposed to the bottom. Also, they were selected faster at the top in comparison to females in the same vertical position. Further, in Study 7, I found that participants demonstrated more surprise, as indicated by increased pupil size, when they were presented with female names at the top as opposed to males at the top. They also showed increased cognitive effort when selecting males as powerful when their names appeared at the bottom versus the top. Across the first six studies, I did not obtain any evidence for spatial processing when participants selected females as powerful or powerless and males as the powerless.

In the present chapter, I aimed at exploring the precise mechanisms driving spatial associations in stereotypic thinking about power and gender and whether such spatial associations can be detected across different task conditions.

Studies 8 and 9: The role of vertical positions in spatial simulations

To explore the role of vertical positions in simulations of gender and power, I conducted Studies 8 and 9. Specifically, I wanted to examine whether verticality cues per se might guide people's choices of males and females as powerful or powerless. To do that, I aimed at excluding the salience of gender by presenting participants with the same gendered names in each trial, that is, two female names and two male names, each one of them in each vertical position (top and bottom). To further investigate the impact of constraining the context of thinking about power and gender to social status, I matched gendered names with high-status and gender-neutral professions in Study 8, whilst in Study 9, I presented gendered names only. Overall, I aimed at measuring how quickly participants would respond to choices of vertical positions when indicating powerful or powerless gendered names. That is, in the current studies, the *position choice* was the quasi-experimental variable.

As Schubert (2005) showed that power is associated with verticality, and my previous experiments demonstrated that verticality indeed plays a role in facilitating people's judgements about power and gender, I explored whether participants would be significantly quicker at categorising individuals as powerful at the top than bottom of the screen across both blocks. That is, in blocks where pairs of male names (block 1) and female names (block 2) were presented. It is likely that under ambiguous conditions whereby only names are presented, participants would mainly rely on spatial locations to make a judgement about power or lack of power of presented names. Such results would indicate that power is associated with verticality even when no strong power judgements can be made about presented individuals. It is important to note that such conditions are different to Schubert's (2005) studies, as gendered names, as opposed to generic powerful and powerless groups (e.g., master versus servant), do not have strong power implications, unless they are thought about in stereotype-consistent ways (like in my previous studies). Power is a relational concept (see Guinote, 2017, for an overview), therefore, it needs to be compared between individuals in a meaningful way. It is therefore less plausible that people would reason about males as more powerful than females, when they do not directly compare them. Hence, it might be easier to capture the effects of verticality cues in the representation of gender within conditions whereby strong power differences are not present (i.e., comparing the same-gender names). To summarise, in Studies 8 and 9, I wanted to test whether power-verticality association is detectable whenever people attribute power or lack thereof to a target individual in the absence of strong power differences between that target considered as powerful/powerless and another individual. I predicted that if power-verticality associations are stable across contexts, then participants should be quicker at indicating top vertical positions when selecting names as powerful due to the top-power association (Schubert, 2005). However, in the light of my previous studies, I did not predict any simulations of names as powerless.

This pattern should be noticed across male- and female-names pairs, as gender should be less relevant in this context. Because participants would not compare power between females and males directly, and power is a relative concept, I predicted smaller differences between categorisations of males as powerful when their names appeared at the top versus females as powerful when their names appeared at the top. I further explored whether such effects would be especially pronounced in the case of Study 8, as additional cues about high social status (i.e., professions) could lead to higher salience of power and make the context more power-relevant. Pairing high-social-status professions with gendered names could provide less ambiguity in the task and stimulate participants to think about more meaningful power differences than when presented with names only. As detected in Study 3 – only high-status professions were judged to be gender-neutral, and so I used the same high-status professions in Study 8. Therefore, I predicted that such stimuli would further be associated with a simulation at the top of targets in the powerful condition. This is because high social status was also found to be associated with top vertical simulations (von Hecker et al., 2013).

Overall, these studies might shed more light onto understanding whether verticality cues (i.e., top and bottom positions) themselves contribute to or inform people's judgements about social power independently of other processes (e.g., stereotypic associations). On the one hand, it is possible that any abstract concept that becomes associated with power might be simulated spatially at the top. As suggested by Barsalou (2008), spatial associations support abstract thinking to produce more concrete mental representations in order to assist cognitive processes, such as social judgement. Specifically, space could provide a more vivid representation of an abstract concept (i.e., power) under ambiguous conditions. This process is similar to another form of simulation, namely mental imagery, whereby people mentally represent multi-sensory information. It was found that when such mental imagery is used, it can facilitate problem-solving (Shaver, Pierson, & Lang, 1974; or see MacInnis & Price, 1987, for a review). However, the difference between spatial simulations as seen in the embodiment literature and mental imagery, is that mental imagery is a deliberate process. Instead, embodied spatial simulations are activated automatically when a relevant concept is processed (Barsalou, 2008). Therefore, it is possible that spatial features of power are automatically activated to facilitate judgements when no other diagnostic information is available. On the other hand, spatial simulations might be used in circumstances that are more meaningful. That is, only semantically valid mental representations based on consistent associations or blending of abstract concepts (such as representing a boss or males as powerful) could be linked with spatial features – as there could be no function to simulate abstract concepts that do not carry meaningful content. I examined these issues in Studies 8 and 9.

Method (Studies 8 and 9)

Participants

I recruited 65 Cardiff University undergraduates for Study 8 (the approximate sample sizes were derived from my previous studies, see Study 5 or 6, mean age =

18.68 years; 56 females). Cardiff University undergraduates were also recruited to Study 9 (N = 64; mean age = 18.36 years; 57 females).

Across both studies, I randomly assigned participants to either the powerful or powerless condition. Each experiment lasted 10 minutes.

Materials and design

Spatial task. I used the same basic design in both studies. I manipulated two types of blocks. In one type, I displayed a block of ten trials with only female names (see Appendix 7 for all the names used). In each trial, there was one pair of female names. One of the names was presented at the top of the screen, the other name appeared at the bottom. In each block, participants saw five pairs of names. First, the names were randomly paired with each other and then each name was always presented with the initially paired other name. Each pair was presented twice in one block, such that the same names would appear once at the top and then at the bottom. The order of trials was random. Each block was repeated 4 times, so there were 40 trials within one type of block.

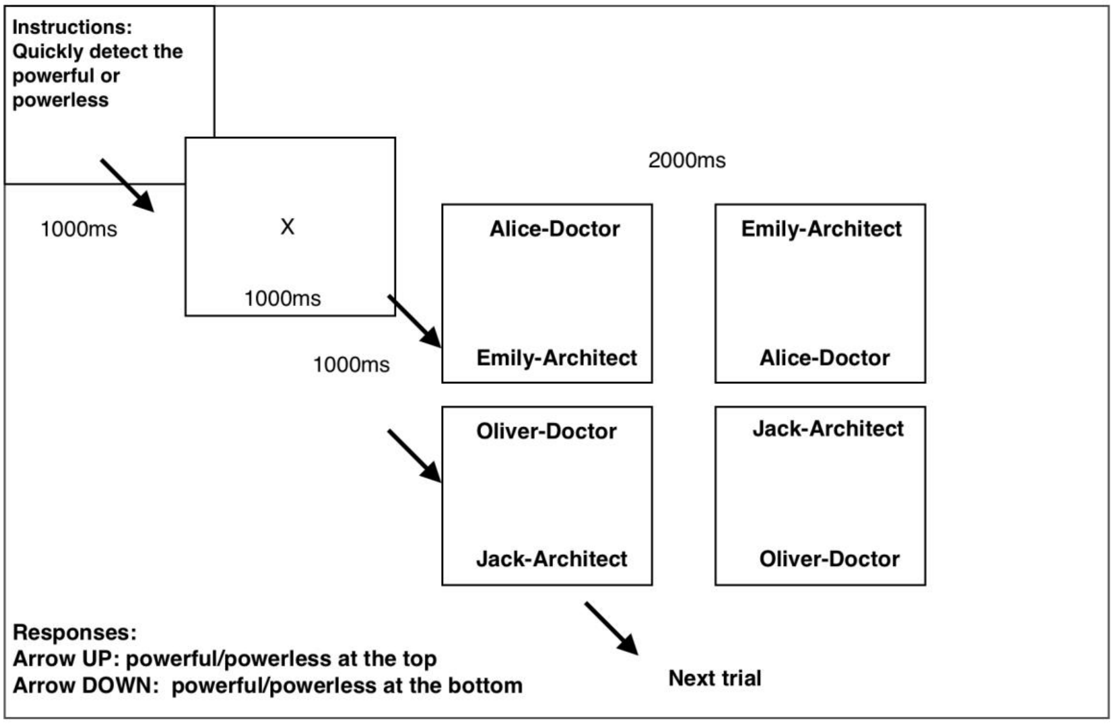
In the second type of block, the design was the same, however, instead of female names, participants were presented with male names.[20] I counterbalanced the order of the type of blocks across participants, such that some participants saw the female blocks first and then the male blocks, whilst the others saw the opposite order. The type of blocks was manipulated within-participants, that is, each participant responded to 80 trials in total. Across these conditions, I asked participants to either quickly select the powerful or the powerless person on the screen – this was manipulated between-participants. In terms of stimuli, I presented the same matched pairs of names in both studies, however, in Study 8, I additionally matched the names with gender-neutral and

[20] I selected the most popular British babies' names (ten male and ten female names) from the same source as cited in Studies 1 - 6.

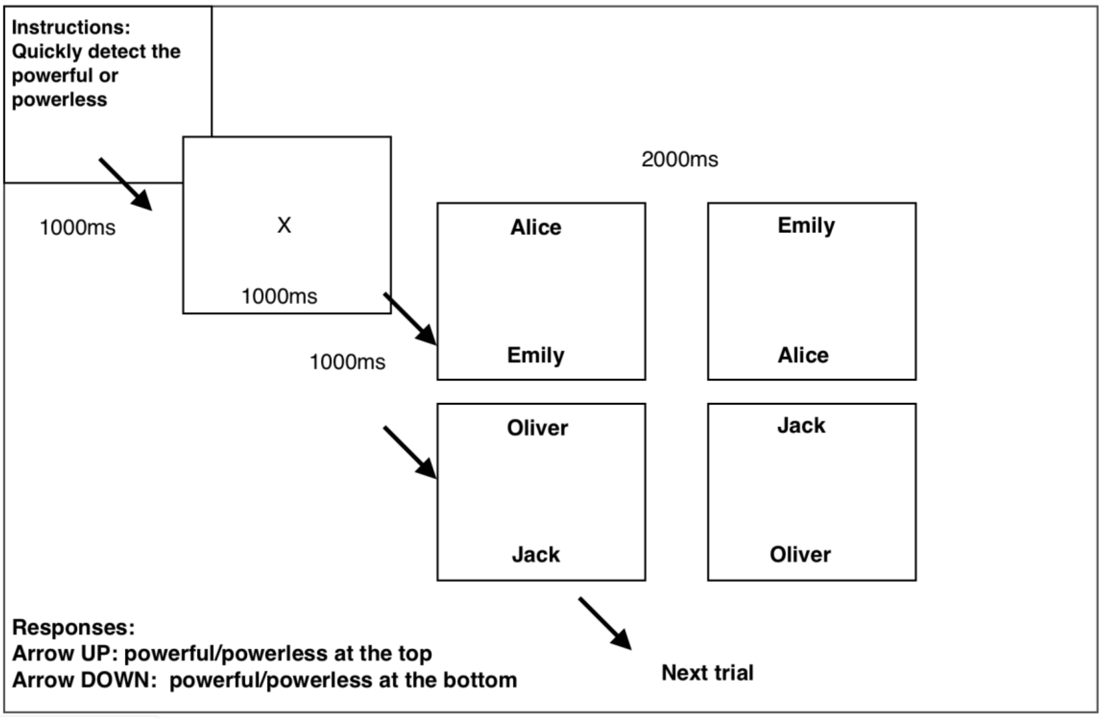
high-status professions (see Figure 13). This was done by adapting the professions from Study 3. Each name was combined with each profession across four different sets of combinations (see Appendix 7 for all the matched pairs of names and professions and their spatial positioning on the screen, Studies 8 and 9). Across both studies I measured participants' reaction times to their choices of top and bottom positions (i.e., the choices of position was a quasi-experimental variable) and also the frequency of these choices in each block and condition.

<u>Figure 13</u>. Spatial task procedure of a single trial

(a) Study 8



(b) Study 9



Procedure

First, participants were instructed to quickly categorise the powerful or the powerless person (depending on the task condition). As the same gender names might be confusing as to which name was powerful or powerless, I told participants that they would be presented with fictitious people and they should try to spontaneously imagine that one of the individuals was powerful (or powerless) and only then make their responses. I also provided a written and verbal definition of power (see Procedure of Studies 1 and 2 for details).

Results: Study 8

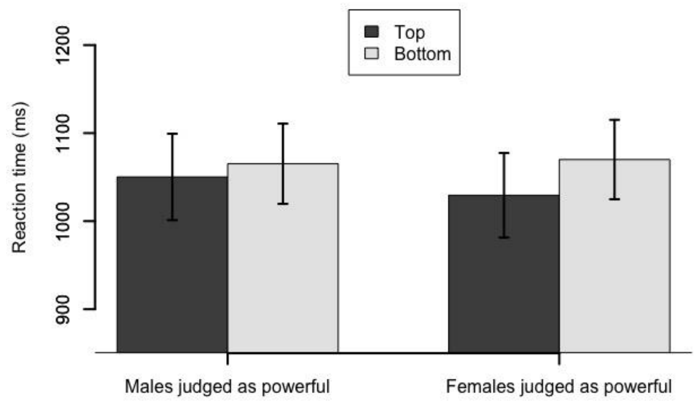
Proportion of choices. First, I analysed how often participants picked powerful and powerless individuals at the top and bottom of the screen. Overall, I found that across both blocks of trials (involving male and female names), participants were marginally more likely to pick names as powerful at the top of the screen (53.2%) than bottom (46.8%), and the same was true for names selected as powerless (top: 50.6% versus bottom: 49.4%), $\chi^2(1, N = 65) = 3.55$, $p = .06$, $d = .48$; odds ratio $_{\text{choosing powerful on top versus bottom}}$ = 1.11 (95%CI[1.00,1.24]) and odds ratio $_{\text{choosing powerless on bottom versus top}}$ = .95 (95%CI[.90, 1.01. Further, on trials with male name pairs, the top name (53.7%) was marginally chosen more often than the bottom name (46.3%) in the powerful condition; I observed the same pattern for male name pairs in the powerless condition (top name: 50.3% versus bottom: 49.7%); $\chi^2(1, N = 65) = 2.82$, $p = .09$, $d = .43$; odds ratio $_{\text{choosing males as powerful on top versus bottom}}$ = 1.14 (95%CI[.98,1.34]) and odds ratio $_{\text{choosing males as powerless on bottom versus top}}$ = .93 (95%CI[.86, 1.01]). In the case of female name pairs, participants were equally likely to pick the top (52.8%) and bottom names (47.2%) in the powerful condition; the same was true for female pairs in the powerless condition, that is, the names that appeared at the top (50.8%) were chosen as often as the names presented at the bottom (49.2%), $\chi^2(1, N = 65) = .97$, $p = .33$, $d = .25$; odds ratio $_{\text{choosing females as powerful on top versus bottom}} = 1.08$ (95%CI[.93,1.27]) and odds ratio $_{\text{choosing females as powerless on bottom versus top}} = .96$ (95%CI[.89, 1.04]). The same results were observed when each condition (powerful and powerless) were analysed separately. Also, the interaction among task, gender block, and position tested in a loglinear analysis was not significant, $\chi^2(1, N = 65) = 4.20$, $p = .65$, $d = .53$, indicating equal distribution of position choices across all the conditions.

Response latencies. To analyse participants' reaction times to their choices, I estimated a linear mixed model. I found that the fixed effects of gender block, $F(1, 57) = 1.98$, $p = .16$, $dz = .06$, and position choice, $F(1, 4657) = .01$ $p = .99$, $dz = .04$, were not significant. However, the fixed effect of task was significant, $F(1, 63) = 13.31$, $p < .001$, $dz = .64$, indicating that participants were significantly quicker at selecting powerful individuals ($M = 1053$, $SE = 57.77$, 95%CI[952, 1179]) than powerless ones ($M = 1346$ms, $SE = 56.89$, 95%CI[1250, 1473]).

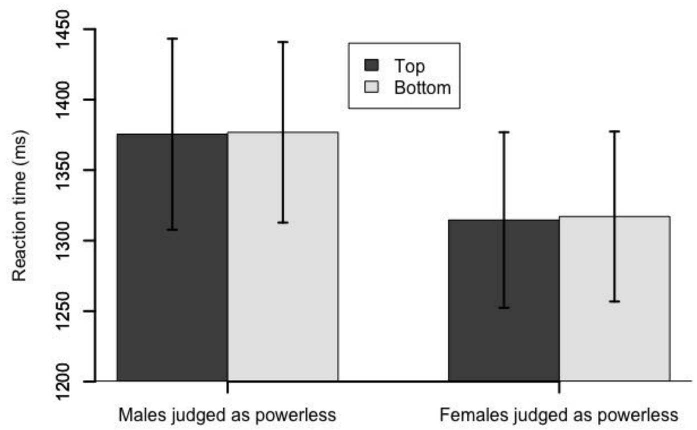
The two-way interactions between gender block and position choice, and gender block by task were not significant, $ps > .81$, and the three-way interaction among gender block, position choice, and task was also not significant, $F(1, 4667) = .03$, $p = .87$ (see Figure 14). However, the position choice by task interaction was marginally significant, $F(1, 59) = 3.31$, $p = .07$. I found that participants were significantly faster to categorise all names presented at the top in the powerful condition, ($M = 1040$ms, $SE = 58.11$, 95%CI[943, 1171]) versus powerless, ($M = 1344$ms, $SE = 57.27$, 95%CI[1258, 1483]), $t(65) = -3.84$, $p < .001$, $dz = .68$. The same was pattern was true for the bottom positions – participants were significantly faster to select all names when they appeared at the bottom in the powerful condition, ($M = 1068$ms, $SE = 58.22$, 95%CI[961, 1189]), than the powerless one, ($M = 1347$ms, $SE = 57.28$, 95%CI[1241, 1466]), $t(65) = -3.41$, $p < .01$, $dz = .64$. In contrast to the hypotheses, in the powerful condition, the chosen names were not selected faster when they appeared at the top than bottom, $t(4666) = -1.28$, $p = .57$, $dz = .23$, and in the powerless condition, the chosen names appearing at the bottom were not selected faster than the ones at the top, $t(4656) = 1.29$, $p = .57$, $dz = .02$.

Figure 14. Response latencies to blocks with (a) male and female pairs in the powerful condition, and (b) male and female pairs in the powerless condition according to the choices of top versus bottom positions. The error bars show -/+1 standard error.

(a) Powerful condition



(b) Powerless condition



Results: Study 9

Proportion of choices. Overall, across both blocks involving the presentation of male and female pairs, I found that participants in the powerful condition were more likely to select names that appeared at the top of the screen (54.1%) versus the bottom (45.9%). A complementary effect was found in the powerless condition, with the bottom names more likely to be selected (50.2%) than the top names (49.8), $\chi^2(1, N = 64) = 9.22$, $p < .01$, $d = .85$; odds ratio $_{\text{choosing powerful on top versus bottom}}$ = 1.20 (95%CI[1.07,1.34]) and odds ratio $_{\text{choosing powerless on top versus bottom}}$ = .91 (95%CI[.82, .97]). Interestingly, I observed the same pattern of results in the block involving male pairs. There was a higher a tendency to pick males as powerful at the top (53.3%) versus bottom (46.7%), and vice versa, however, with a smaller magnitude, for the males as powerless at the bottom (51.0%) versus top (49.0%), $\chi^2(1, N = 64) = 8.59$, $p < .01$, $d = .78$; odds ratio $_{\text{choosing males as powerful on top versus bottom}}$ = 1.27, 95%CI[1.08,1.48]) and odds ratio $_{\text{choosing males as powerless on top versus bottom}}$ = .88, 95%CI[.81, .96]). However, that was not the case when female targets were chosen, as participants picked them equally likely at the top and bottom across both conditions,[21] $\chi^2(1, N = 64) = 2.30$, $p = .13$, $d = .38$.[22] Such results may indicate that associations between power and verticality were more pronounced in the case of thinking about males as opposed to females, even in the stereotype-inconsistent choice (powerless-males).

Response latencies. In terms of reaction times, I estimated a linear mixed model and I found that the fixed effect of gender block was significant, $F(1, 62) = 7.37$, $p < .01$, $dz = .20$, which indicated that participants were faster in the blocks involving female targets (M = 1002ms, SE = 24.61, 95%CI[951,1047]), as opposed to male targets

[21] Females as powerful at the top (53.3%) versus bottom (46.7%) and females as powerless at the top (49.4%) versus bottom (50.6%). Odds ratio $_{\text{choosing females as powerful on top versus bottom}}$ = 1.05, 95%CI[.95,1.31]) and odds ratio $_{\text{choosing females as powerless on top versus bottom}}$ = .94, 95%CI[.87, 1.03]).

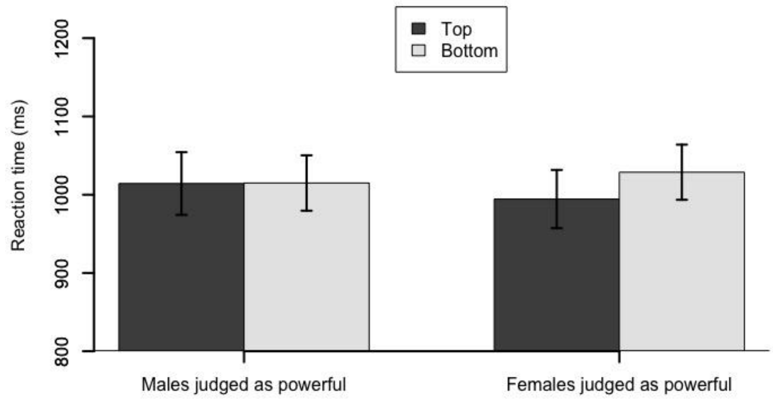
[22] These results were also confirmed by a loglinear analysis that indicated a significant interaction among task, gender block, and position, $\chi^2(1, N = 64) = 10.50$, $p < .03$, $d = .89$.

(*M* = 1041ms, *SE* = 29.76, 95%CI[984, 1101]). The fixed effect of position choice, *F*(1, 4811) = 2.71, *p* = .10, *dz* = .05, and task, *F*(1, 62) = .26, *p* = .61, *dz* = .03, were not significant.

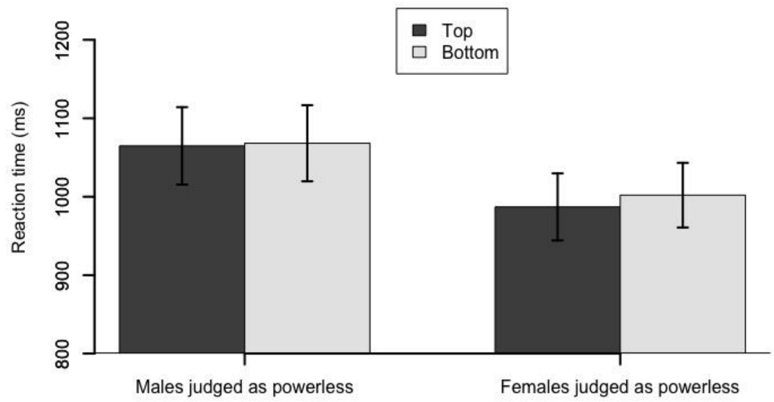
In terms of two-way interactions, I found that gender block by position choice as well as position choice by task interactions were not significant, *ps* > .11. The task by gender block interaction was marginally significant, *F*(1, 62) = 3.35, *p* = .07. The post-hoc comparisons indicated that participants were faster to categorise female targets as powerless, (*M* = 994ms, *SE* = 34.78, 95%CI[928, 1067]), than male targets as powerless, (*M* = 1019ms, *SE* = 42.09, 95%CI[986, 1154]), *t*(62) = 3.22, *p* < .01, *dz* = .14., but there were no differences in the speed of choices of males as powerful *(M* = 1015ms, *SE* = 42.10, 95%CI[930, 1098]) versus females as powerful, (*M* = 1010ms, *SE* = 34.81, 95%CI[931, 1070]), *t*(62) = .63, *p* = .92, *dz* =.03. Other comparisons were not significant, *ps* > .58. As in Study 8, the predicted three-way interaction (gender block by position choice by task) was not significant (see Figure 15).

Figure 15. Response latencies to blocks with (a) male and female pairs in the powerful condition, and (b) male and female pairs in the powerless condition according to the choices of top versus bottom positions. The error bars show -/+1 standard error.

(a) Powerful condition



(b) Powerless condition



Summary of results: Studies 8 & 9

The aim of these studies were to test whether verticality cues (top and bottom positions) were used when thinking about power differences between two targets that did not carry strong power implications (i.e., pairs of male or female names). The results of the studies imply that this was not the case, as selecting target names as powerful at the top did not facilitate participants' responses. To summarise, spatial cues do not seem to contribute to spatial simulations, when they are not presented in a meaningful power-related context. My results further indicated that participants were quicker to detect powerful versus powerless individuals at the top, but this was also true for the bottom positions. This happened only when the subtle power context was activated, i.e., when gendered stimuli were presented with high-status professions in Study 8. The same was not true in Study 9, where the main finding was that participants tended to be quicker in categorising female targets as powerless than male targets as powerless. However, I found no complementary findings in the powerful condition.

In terms of frequency of top and bottom choices, in Study 8, participants were marginally more likely to select top positions within both conditions, that is, when choosing targets as powerful and powerless. In Study 9, however, in the powerful condition, top positions were selected more often when indicating male targets as opposed to the bottom. Also, in the powerless condition, male targets were selected more often at the bottom as opposed to the top. Although when collapsing across gender blocks, this was true for both targets, the separate analyses indicated that associations between powerful-top and powerless-bottom were only true for male pairs. This was also supported by a significant interaction among position, gender block, and task. These findings, however, are only confined to the frequency of choices, and not the spatial simulations as investigated by measuring response latencies of those choices.

This suggests that participants used spatial cues only to *decide* who was powerful/powerless when choosing males. That is, spatial cues were only associated with the *outcome*, but not the *duration* with which participants processed the cues.

Overall, it is possible that spatial simulations do not occur when the task conditions are too ambiguous and not meaningful to participants. To investigate the role of spatial cues in gender categorisations under unambiguous conditions, but in the absence of thinking about power, I conducted Study 10.

Study 10: Gender simulation

Across seven studies (Studies 1 - 7), I have established that stereotypical thinking about power and gender involves a pronounced spatial simulation of males as powerful at the top. So far, I investigated such effects when both gender and power concepts were made salient to participants. In the current study, I first aimed at testing whether thinking about gender per se can be associated with verticality without a salient power context. It is plausible that not only does thinking about power as well as stereotypic associations involve mapping of social concepts onto the vertical spatial dimension, but also gender itself might contribute to such representations. In this study, my first aim was to explore whether the previously detected spatial effects (i.e., males simulated at the top as powerful) were also linked to an independent association between gender and space. Specifically, I wanted to examine whether the concept of gender per se is also grounded in space.

Second, it is important to consider the potential relevance of polarity correspondence theory. Polarity correspondence theory, suggested by Proctor and Cho (2006), postulates that linguistic unmarked ends of a dimension (e.g., up or right) are +polar (i.e., they have positive polarities), whilst marked ends (e.g., down or bottom) are –polar (i.e., they have negative polarities). Then, matching polarities between response modes and vertical positions of stimuli produces faster responses in reaction-time tests. Specifically, if polarities of spatial locations of stimuli (e.g., top or bottom) are matched with the polarities of the response mode (right or left, respectively), then a facilitation in reaction-times to such stimuli would occur (see Proctor & Xiong, 2015; Weeks & Proctor, 1990). Further, such faster responses are associated *only* with the structural overlap of polarities, rather than semantics of the concepts. Therefore, in the current study, I also wanted to explore whether response mode (i.e., categorising gendered names at the top with a right key on the keyboard versus left key) would be associated with faster responses in line with the polarity correspondence theory.

Overall, in the present study, I explored whether (a) gender per se is semantically associated with the spatial vertical dimension without a power relevant context, and (b) the structural overlap between locations of stimuli and response modes contributes to the speed of participants' reaction times in spatial tasks used in my previous studies. To test such possibilities, I asked participants to quickly categorise gendered names as either female or male. The names were presented vertically on the screen (top or bottom). Participants saw one name at the time. I measured their response latencies when they categorised the names in each vertical position. I also manipulated the response mode. Half of the participants responded with a left key (A) to indicate male names and a right key (L) to indicate female names, whilst the other half responded with L to male names and A to female names on the computer keyboard.

First, if gender is associated with the spatial vertical dimension and does not require power judgements to be spatially simulated, then I predicted that participants would be quicker at categorising male names as *male* when they appeared at the top of the screen versus bottom. Second, if the structural overlap between response modes and locations per se is responsible for processing facilitation, then participants should be faster at categorising both male and female names when they appeared at the top of the screen when using the right key. They should also be faster at responding to male and female names when they appeared at the bottom and when using the left key. Finally, based on the previous literature indicating the role of top-down reading habits (e.g., Meier & Robinson, 2004; Schubert 2005; von Hecker et al., 2013; von Hecker et al. 2016), I predicted that regardless of gender and response modes, participants should be faster at sorting the names when they appeared at the top as opposed to the bottom. Faster responses to top stimuli would suggest that participants scanned the screen from the top to the bottom, and therefore were more likely to attend to the top stimuli first and bottom stimuli as second.

Method

Participants

I recruited 60 participants (I derived the sample size from my previous experiments: Study 2 for Cohen's *dz* = .25-.35; mean age = 18.25, 52 females), but I excluded one participant, as their data were missing, whilst a second was too slow at responding according to the Tukey criterion. I randomly assigned participants to press either the letter A on the computer keyboard to categorise male names and letter L to categorise female ones, or A to categorise females and L to categorise males. They received course credits for their participation.

Materials

I selected ten of the most popular British babies' names in 2015,[23] five of which were female (Amelia, Olivia, Jessica, Daisy, and Phoebe) and five were male (George, Oscar, Alexander, William, and Joseph).

[23] See: *The most popular baby names.* (n.d.). Retrieved August 15, 2016, from: http://www.mumsnet.com/baby-names/most-popular-baby-names-england-wales.

Design

I used a mixed design. The response key condition (A-male or A-female) was manipulated between-participants, whilst gender of the stimuli presented (male or female names) and their position on the screen (top or bottom) were manipulated within-participants. There were four blocks of trials in total and in each block participants saw each male and female name (one at a time) presented twice, once at the top and once at the bottom of the computer screen. That is, there were 20 trials in total in each block. The order of trials was randomised.

Procedure

Participants were informed that they would see a single name on the screen in each trial, either at the top or at the bottom of the screen and that their task would be to quickly and accurately categorise the displayed name as male or female. They were instructed to use index fingers. Each trial began with a display of a blank screen for 1000ms, which was followed by a fixation mark X in the middle of the screen presented for 1000ms as well (participants were instructed to look at the fixation). Then a stimulus name was presented. The names were centred horizontally and displayed at the top or at the bottom of the screen. Before the main task, participants were given five practice trials. In each trial, a name appeared for 2000ms, if participants did not make a response, the name would be cleared from the screen. A response was required to proceed to the next trial, even if the stimulus disappeared. If they made a mistake, the word INCORRECT appeared on the screen (in red colour). To proceed to a next trial, participants needed to correct their responses. The names were presented in white letters on a black screen (font size 17). After each block participants could take a break. The experiment lasted approximately 10 minutes.

Study 10: Results and discussion

First, I removed the incorrect responses (this happened when participants categorised male names as female or vice versa; these trials accounted for 3% of all responses). Subsequently, as in my previously reported studies, I estimated a linear mixed model to analyse response latencies. The fixed effect of gender was not significant, $F(1, 56) = .08$, $p = .78$, $dz = .03$, but I found that the effect of position was significant, $F(1, 330) = 11.36$, $p < .001$. $dz = .30$. Participants were significantly faster at categorising names when they appeared at the top of the screen (M = 693ms, SE = 11.60, 95%CI[672, 719]) as opposed to the bottom (M = 710ms, SE = 11.74, 95%CI[687, 736]). The fixed effect of response mode was also significant, $F(1, 350) = 10.43$, $p < .001$, $dz = .30$, such that participants were significantly faster to respond in the condition where they categorised males with L and females with A (M = 694ms, SE = 13.87, 95%CI[651, 707]), as opposed to the condition where they responded to males with A and females with L (M = 710ms, SE = 13.44, 95%CI[700, 756]).

Further, the two-way interactions (gender choice x vertical position, see Figure 16; gender choice x response mode; and vertical position x response mode), were not significant, $ps > .18$, as was the three-way interaction between response mode, gender choice, and vertical position, $F(1, 3961) = 2.57$, $p = .11$ (see Figure 17). These results indicate that gender per se is not associated with verticality. When participants sort gendered names as male or female without the power context they do not use spatial cues to facilitate their judgment. This is true, even though participants exhibited facilitation to stimuli presented at the top, as this is likely associated with reading habits (see von Hecker et al., 2016).

Figure 16. Response latencies as a function of presented gendered names and their vertical position. The error bars show -/+1 standard error.

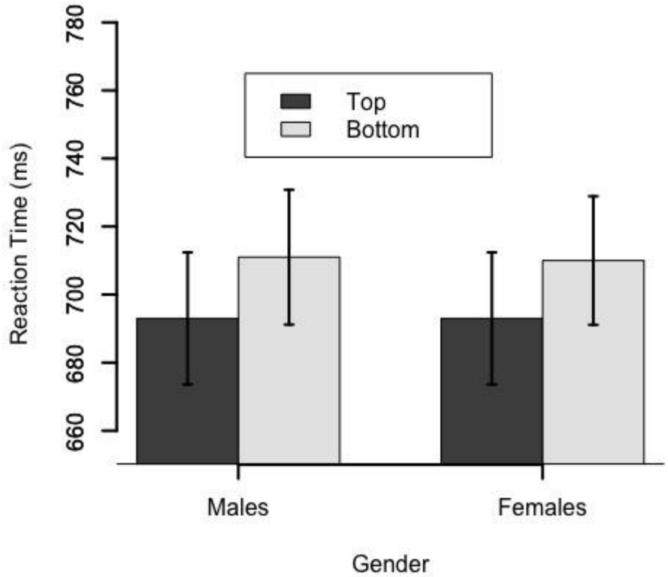
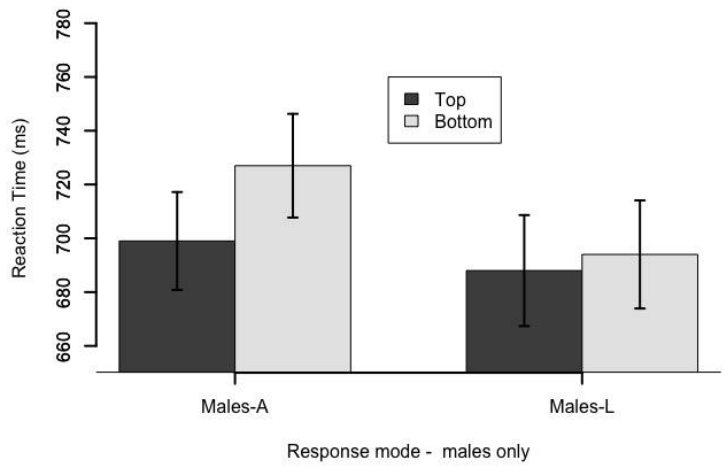
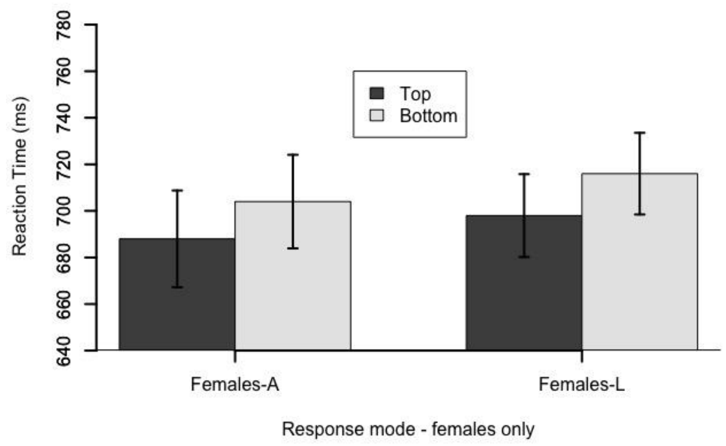


Figure 17. Response latencies as a function of gendered names, their position, and response mode. The error bars show -/+1 standard error.

(a) Categorisations of males



(b) Categorisations of females



Summary of results

To summarise, in this chapter, I aimed at investigating whether spatial simulations occur when task conditions are modified such that power or gender cues become less salient. The results from Studies 8, 9, and 10 indicate that in general spatial simulations do not occur in the absence of meaningful power differences between targets presented or when the power-context is not salient. Only in Study 9 were spatial cues associated with responding in the male gender block and both powerful and powerless conditions (top choices were associated with categorisations of targets as powerful and bottom choices with categorisations of targets as powerless), but not in the female block. However, this was not observed in Study 8. Also, verticality cues did not involve response facilitation across both studies (8 and 9). Similarly, in Study 10, I found that spatial cues did not facilitate participants' categorisations of gendered names. Specifically, gender per se was not associated with verticality, as the two-way interaction between gender (males versus females) and position was not significant. Participants were equally fast at categorising female and male names at the bottom and top of the screen. Further, participants were faster to categorise both male and female names at the top of the screen versus bottom, possibly suggesting that the reading habit (i.e., scanning the screen from the top to the bottom) was responsible for such a processing advantage (von Hecker et al., 2016). Also, there was a processing advantage in the condition where participants responded with L to male names and A to female names in contrast to the condition where male names were categorised with A and females names with L. This is inconsistent with polarity correspondence theory that I discuss in more detail in Chapter 7. However, it is important to note that these results could be related to handedness of participants – which I did not control for in the present study.

To conclude, in the light of my previous studies, it appears that stereotype accessibility in the power context is one of the contextual factors that is required for spatial simulations to emerge. Therefore, in the next chapter, I investigated to what extent such stereotype activation is important for people to simulate power and gender on the vertical dimension.

Chapter 5: Dependency of spatial simulations on stereotype accessibility

Overview

In the present chapter, I investigated to what extent spatial simulations in the abstract concept of power are associated with stereotype accessibility. In two studies, I asked participants to complete a spatial task where they were instructed to categorise as quickly as possible which person (between two vertically presented on the screen, one male, one female) was powerful. However, in Study 11, I manipulated stereotype accessibility by restricting the time to complete the task in each trial. In one condition, participants received just 600ms to select the powerful person (short condition), whilst in the other they had 2000ms (long condition). I measured how often participants selected each gendered name in each vertical position (top versus bottom) across the time-constrained conditions. In Study 12, I attempted to reduce stereotypic associations by adapting the evaluative conditioning task (Olson & Fazio; 2001, 2006). In this task, participants were implicitly primed with pictures of females paired with powerful concepts and males paired with powerless concepts. Then, they completed the spatial task and categorised powerful persons. In study 11, I found that participants relied on the gender stereotype (males-powerful) more in the long versus short condition, but across both conditions they were more likely to associate both male and female names with top positions versus bottom positions. These results imply that when people make power judgements, they base their decisions on verticality cues regardless of the activated stereotype. As I did not measure the *duration* of participants' decisions (RTs), these results can only suggest that the outcome of decision was based on verticality cues. That is, these results cannot indicate whether spatial simulations occurred. In Study 12, I did not detect any evidence of spatial processing across both groups of participants, possibly to a significant data loss.

In the previous chapters, I found that mental representations of gender and power involve spatial features associated with vertical positions. Males as powerful seem to be simulated at the top of the vertical dimension as opposed to females as powerful. I found that such spatial features can be detected only in certain contexts. Spatial simulations take place when people reason about gender and power in a stereotype-consistent manner and only when they think about powerful as opposed to powerless individuals. However, spatial simulations do not occur when the concepts of power and gender are not simultaneously salient in the experimental conditions. Therefore, to date, I explored the nature of spatial simulations under different experimental conditions by investigating the role of gender in power-relevant context. In the current chapter, I present two studies that directly explored the role of stereotype activation in the spatial simulation of gender and power on the vertical dimension. As it appears that spatial simulations occur when stereotype-fit is present, it is possible that manipulating the accessibility of a stereotype can also affect the degree to which spatial simulations would be present. I addressed these issues below.

Study 11: Stereotype activation under restricted time conditions[24]

First, in Study 11, I tested whether gender has power implications and under what circumstances such power implications might emerge. I adapted Schubert's (2005) paradigm and presented participants with vertically arranged pairs of gendered names. On some trials, a male name appeared at the top of the screen and a female name at the bottom, or vice versa. I instructed participants to identify the powerful person of the pair. I predicted that participants would be more likely to pick males as powerful in contrast to females. This association was expected to be moderated by the vertical position of stimuli – both males and females should be seen as powerful more

[24] The report of Study 11 is partly based on a manuscript in preparation (Zarzeczna et al., 2018b)

often at the top as opposed to the bottom due to the link between power and verticality (Lakoff & Johnson, 1999; Rudman & Kilianski, 2000; Schubert, 2005). Yet, if males-top associations are stronger than females-top via stereotype-fit, participants should select more males at the top relative to females. Second, I investigated whether the identification of males and females as powerful would be associated with time constraints. The empirical literature indicates that stereotype activation is dependent on the thought context (Fiske & Taylor, 2010). When categorising individuals, people attend to and rely on information that is more immediately relevant and accessible to the task (Macrae et al., 1997). Therefore, in this study I varied the amount of time in which the stimuli were presented. I hypothesised that in shorter trials (stimuli presented for 600ms), participants would rely more on salient verticality cues (top/bottom positions) that are more immediately linked with power than gender (Schubert, 2005) and pick more individuals at the top regardless of gender. With longer presentation intervals, I expected that participants would make more stereotype-consistent choices, selecting more males as powerful than females.

Method

Participants

I recruited 62 Cardiff University Psychology students (56 females, mean age = 19.94 years). The sample size in both studies was derived from von Hecker et al. (2013, 2016) for Cohen's $dz = .25$. They were randomly allocated to the short or long condition and gained course credits for their participation.

Materials

I selected 20 popular British names (10 female, 10 male). I subsequently paired each name with a name of the opposite gender (e.g., Ethan – Zoe, see Appendix 8 for all the pairs). The pairs were then presented on a computer screen with presentation software DirectRT (Jarvis, 2012) in white letters on a black screen (font size 15 – 21 pixels).

Design

I manipulated the vertical position of the paired names within-participants, such that on half of the trials, a male name (e.g., Ethan) was displayed at the top of the screen with a female name (e.g., Zoe) displayed the bottom. On the other half of the trials, this display was reversed (see Figure 18). The time restriction for participants' judgements about the powerful person was manipulated between-participants. In the long condition, the stimuli were presented for 2000ms, as in my previous experiments; in the short condition the stimuli were presented for 600ms, which allows for only a quick judgement. I measured the frequency of participants' choices of each gender as powerful in each vertical position. There were four blocks of trials in both conditions. Each block involved a presentation of 10 gendered pairs shown twice, with each name presented in each vertical position (20 trials in total).

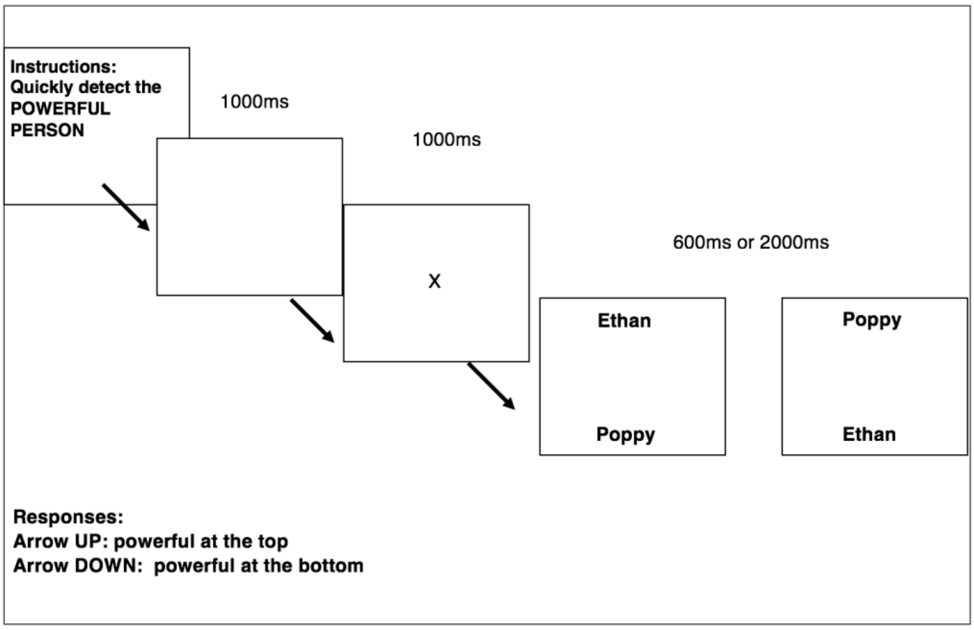
Both names were centred and there was a 23cm vertical distance between them. There were 80 trials in total. The order of trials within each block was randomized. Participants sat approximately 70cm away from the computer screen.

Procedure

First, I presented participants with task instructions on a computer screen. They were informed that they would be presented with names of fictitious people and in each trial they would see two names. They were then informed that these names would be presented for 2000ms/600ms (depending on experimental group) and that they could respond even after the stimuli disappeared from the screen. Thus, the response interval was open-ended, and participants were not instructed to respond as quickly as possible.

Second, I instructed participants that their task would be to detect the powerful person of the pair displayed on the screen. I emphasised that they should try to spontaneously imagine that one of the people was the powerful one. Third, to ensure that participants made judgements about *social power,* I presented a definition of a socially powerful individual, adapted from Galinsky, et al. (2003), as in previous studies. Finally, participants used the arrow keys on the computer keyboard to indicate the powerful person (*arrow up* to choose top stimuli and *arrow down* to pick bottom stimuli; see Figure 18). The task took approximately 15 minutes.

Figure 18. Spatial task procedure of a single trial.



Results and discussion

To analyse the proportions of participants' choices of males and females as powerful in each vertical position, I conducted a loglinear analysis.[25] First, I found a main effect of gender. Participants were more likely to select male names as powerful than female names, $\chi^2(1, N = 62) = 457.27$, $p < .001$, $d > 2.0$ (odds $_{\text{choosing males over females}}$ = 1.87).[26]

[25] Due to the underrepresentation of males in my sample, I weighted cells within the loglinear analysis by participants' gender. Subsequently, I ran the same analysis without male participants. I obtained the same results across both analyses when excluding and including male participants.
[26] This effect size was too large to obtain the exact value of Cohen's d.

Second, as predicted, I found the main effect of condition in relation to gender.[27] Participants were more likely to select male names as powerful in the long condition (68.8%) than in the short condition (61.6%), $\chi^2(1, N = 62) = 27.83$, $p < .001$, $d = 1.80$ (odds ratio $_{\text{choosing males in the long versus short}}$ = 1.37 95%CI[1.22,1.54]).

However, contrary to predictions, the main effect of condition on power allocation to the top vertical position was not significant, $\chi^2(1, N = 62) = 2.27$, $p = .13$, $d = .40$. This demonstrates that participants were not differentially relying on top vertical positions of stimuli in the short (53.2%) versus long condition (51.1%; odds ratio $_{\text{choosing top in short versus long}}$ = 1.04, 95% CI[.99, 1.10])

Finally, I found a significant main effect of position on power allocation, $\chi^2(1, N = 62) = 10.17$, $p < .001$, $d = .89$ (odds ratio $_{\text{choosing top versus bottom}}$ = 1.21, 95%CI[1.08,1.36]). However, this effect was not moderated by gender. Participants were marginally more likely to select the male name as powerful when the male name appeared at the top (67.3%) of the screen compared to the bottom (63%), $\chi^2(1, N = 62) = 3.54$, $p = .06$, $d = .49$. The same pattern was true for selections of the female as more powerful. Participants were more likely to select the female name as more powerful when the female name appeared at the top (37%) of the screen compared to the bottom (32.7%), $\chi^2(1, N = 62) = 6.63$, $p < .01$, $d = .69$.

As hypothesised, in Study 11, I demonstrated that gender has social power implications, such that while thinking about power people conform to gender stereotypes, perceiving males as more likely to be powerful than females (Rudman & Kilianski, 2000). Consistent with my hypothesis, I also found that participants relied on such a stereotype more in the long condition than in the short one. Specifically, more processing time enables people to use stereotypes to guide their responses while picking

[27] I present the results in terms of main effects for clarity, however, chi-square tests indicate associations between variables.

powerful individuals (Fiske & Taylor, 2010). However, in the short condition, participants did not rely more on verticality cues. Also, I found that whenever people considered both males and females as powerful, they were more likely to associate them with top vertical positions (Schubert, 2005). These findings diverge from my initial studies (1-7), which suggested that only males as powerful are simulated at the top. At the same time, in the present study, I measured only *associations* in terms of the frequency of participants' choices, which represents a different methodology to reaction-times or pupillometry. It is likely that the frequency of choices measures mainly the *outcomes* of thought processes, while the reaction-time/pupillometry methodologies focus more on the *duration* of the thought process itself and the cognitive effort associated with that process. For example, participants might try to control their stereotype by deliberately choosing more females as powerful at the top, but their choices would be much slower than choices of males as powerful at the top (Studies 1 - 7). This is because the cognitive effort associated with processing of females as powerful at the top would be higher than in the case of males due to stereotype misfit. This process, however, would not be noticeable in the case of proportions of choices, as observed in the present study. I explored this issue further in Study 12.

Study 12: Reducing stereotypic thinking

Overall, Study 11 provides evidence that associations between power and spatial location are present for both genders and are not moderated by stereotypic associations. Such findings are not in line with my previous studies. However, note that the previous results were based on reaction-time methodology and pupillometry. Therefore, to explore whether stereotypic thinking moderates *spatial simulations*, but not *spatial associations,* of gender, I conducted a follow-up study. In this study, I attempted to experimentally reduce stereotypic-associations. If associating power and gender involves spatial simulations that rely on consistency between concepts (stereotype-fit between males-power), then reducing that consistency or disrupting the blending between males and power should also involve a disruption of spatial simulations of males at the top and possibly lead to more pronounced simulations of females at the top instead.

In order to reduce the stereotype link between males-powerful and females-powerless, I used a priming task. The task was based on the evaluative conditioning technique validated by Olson and Fazio (2001, 2006), which they found reduced implicit prejudice against African Americans (see Hofmann, De Houwer, Perugini, Baeyens, & Crombez, 2010, for a review). In the standard evaluative conditioning task, participants view a number of pictures of Black human faces paired with positive stimuli and White faces paired with negative stimuli on a computer screen. The priming stimuli are randomly presented with a number of filler and distractor items, such as pictures of objects or humans, to make participants unaware of the priming pairings. To keep participants attentive throughout the procedure, their task is to press a key whenever they see a pre-specified target item (e.g., a woman jogging). The aim of the task is to implicitly condition participants to associate positive valence with Black people and negative with White people.

In my study, I used the same priming procedure, however, my priming pairings consisted of pictures of female faces paired with powerful concepts and pictures of male faces paired with powerless concepts. I included filler pictures of different items and asked participants to press a keyboard key every time they saw a neutral stimulus (e.g., a male or female face paired with a neutral adjective). I also designed a control condition, where participants were presented with the same stimuli as in the priming condition, but the gendered pictures and the powerful/powerless concepts were unpaired, but appeared on the screen separately. Then, I asked participants to complete a spatial task adapted from Study 3, but I only used the powerful condition, as I did not detect spatial simulation effects in the powerless condition in my previous studies.

On the basis of my previous studies, I hypothesised that in the control condition, participants would simulate males at the top, that is, respond faster when selecting male names as powerful when they appeared at the top as opposed to either selecting female names as powerful when they appeared the top, or male names when they appeared at the bottom. This effect was expected to be reduced in the priming condition. I suggest that priming participants with females as powerful and males as powerless might create a new thought context, where the powerful-females link is no longer considered as *inconsistent* or *incorrect* within the task. Under such circumstances, thinking about males would not be blended with power and supported by stereotype-fit resulting in weaker spatial simulations at the top.

Method

Participants

I recruited 49 Cardiff University psychology students (the approximate sample size was derived from von Hecker et al., 2013 for Cohen's d = .25-.35), but I excluded 7 participants. One participant guessed the purpose of the study, 5 were aware of the pairings in the priming task (females-powerful; males-powerless), and another participant was too slow in responding in the spatial task according to the Tukey criterion. Therefore, the final sample size consisted of 42 participants (mean age = 19.65, 32 females). I randomly allocated them to either the priming or control condition. They received payment or course credit for their participation.

Materials

Priming Task. First, I randomly selected a set of 10 pictures of females and 10 pictures of males from the database of 70 pictures created by Lundqvist, Flykt, and Ohman (1998). Each picture represented a person with a neutral facial expression (aged from 20 – 30 years). The pictures were standardised such that each individual was presented in the same manner, including the background of the picture, clothes, and distance from the camera.[28] I adjusted the size of the pictures (368 x 500 pixels) to the computer screen (1280 x 1024 pixels). Second, I randomly selected 20 pictures of filler/distractor objects (e.g., cactus, cd, pavement, chain; size: 500 x 500 pixels), such that each object belonged to a different category (e.g., plants, tools), from a set of standardised stimuli developed by Brodeur, Dionne-Dostie, Montreuil, and Lepage (2010). Subsequently, I randomly paired five pictures of female faces with five concepts associated with power (i.e., power, powerful, control, commanding, influential) and five pictures of male faces with five concepts associated with lack of power (i.e., powerless, submissive, passive, vulnerable, subordinate).[29] Then, I paired the rest of the selected gendered pictures (five females, five males) with five neutral adjectives adapted from Rocklage and Fazio (2015), which were rated by their participants to be neutrally valenced and associated with low emotionality (i.e., average, neutral, adequate, okay, and satisfactory). The pictures and the neutral adjectives were paired such that there was always one picture of a male and female matched with the same adjective. Four pairings out of 10 (two female pictures and the paired adjective; two male pictures and the paired adjective) were randomly selected to be the target items that participants needed to categorise during the priming task. Both the pictures of females and males as well as the adjectives constituted targets and were presented in three different ways. Sometimes each adjective and picture appeared on the screen on

[28] More information about the pictures can be found at: http://www.emotionlab.se/resources/kdef
[29] The concepts were adapted from a dictionary of synonyms at: http://www.thesaurus.com

their own and sometimes they were paired together. This was done to make sure that participants paid equal attention to all the labels and pictures. In terms of the filler objects, a set of 15 items was used in the main experimental phase (some pictures were paired with a consistent label, e.g., a picture of a cactus paired with the word cactus, and some with an inconsistent one, e.g., a coconut picture paired with the word bicycle). Some items were only represented by their labels and some only by pictures, whilst five pictures of items were used in the practice phase. The labels of the pictures and the adjectives were presented in black colour (font size 18). All the stimuli were presented on a white background.

Spatial Task. I adapted the spatial task materials from Study 3 (see method section). Specifically, I used gendered names paired with high-status and gender-neutral professions. As there were four sets of combinations between names and professions, participants were randomly allocated to receive one set, as it was done in Studies 3 and 8.

Design

I used a mixed design. I manipulated the priming task (priming vs. control) between-participants. In the priming condition, I presented four blocks of trials. Each block started with a presentation of the target items that participants needed to categorise during the task (i.e., the picture of a single individual who was either male or female and the label of a neutral adjective) chosen from the set of ten pictures of males/females paired with neutral adjectives. The rest of the pictures (i.e., nine) were presented as filler items and participants were not supposed to categorise them. The selected target items within each block were presented six times (two trials presented just the picture, two others presented just the adjective, and the final two presented the picture paired with the adjective). To further obscure the purpose of the task, I also presented 15 filler items (pictures, labels, and pictures paired with labels) and four blank screens to eliminate a possible sense of rhythm of the presentation of stimuli, as recommended by Olson and Fazio (2006). On ten other trials, participants were exposed to the priming pairings, out of which five represented pictures of females paired with power concepts and five pictures of males paired with powerless concepts. Overall, there were 44 trials in total in each block. The order of the trials was random. The same items were presented across all blocks, except for the target items, which represented different individuals.[30] The order of blocks was counterbalanced within participants. Before the main task, I included 24 practice trials that presented only filler objects and targets. Each stimulus appeared in the centre of the computer screen for 1500ms and was preceded by a blank screen presented for 1000ms. In the control condition, participants completed a similar task, except that the priming gendered pictures and power/powerless concepts were presented separately, which prolonged the task for 15 seconds in comparison to the priming task.

In the next stage of the experiment, participants completed my standard spatial task (designed in the same way as in Study 3). I presented vertically positioned gendered names (one at the top; the other at the bottom), that is, the trial-type was manipulated (males-top versus females-top) within-participants. This time, however, I

---

[30] The design was pre-tested in a pilot study. I created two versions of the priming task and manipulated the occurrence frequency of the priming pairings (females-powerful; males-powerless). In the higher frequency condition, I presented four blocks of trials. Within each block, 10 priming pairings were presented, out of which five pictures of females paired with powerful concepts and five pictures of males paired with powerless concepts were presented. The filler items were also randomly dispersed throughout the procedure. Overall, there were 40 trials in total including the filler trials. In the lower frequency condition, I presented 10 priming pairings twice within two blocks of trials. Each block involved 80 trials in total. Participants' task was the same as in the main experiment, but at the end I asked them to estimate the proportion of trials within the task that presented genders paired with power concepts. Participants in the higher frequency condition estimated a slightly higher number of such trials than those in the lower frequency one, but the difference was not significant, $t(24) = 1.18$, $p = .25$, $d = .03$ (I recruited 29 participants, but 4 participants misunderstood the question and were excluded from this analysis). I also asked whether they were aware that females were paired with powerful concepts/males with powerless. I found that in the higher frequency condition, a higher proportion of participants was unaware of the gendered pictures paired with power concepts than in the lower frequency, but the relationship between the condition and awareness was not significant, $\chi^2(1, N = 25) = .20$, $p = .27$, $d = .19$ Because the awareness was lower in the higher frequency condition, I used that version in the main experiment.

asked participants to categorise just the powerful person, as I was interested mainly in reducing the effects of simulations of males at the top. I measured participants' choices of the powerful individuals (quasi-experimental variable) and also response latencies to all possible choices.

Procedure

First, participants completed either the priming or control task. They were instructed to press the spacebar as fast as possible every time they saw a pre-specified target item on the screen. I also informed them that I was interested in attention and rapid responding and encouraged them to carefully attend to all the stimuli on the screen. Second, they completed the spatial task. At the end, I asked participants whether they noticed anything unusual about the way the words and pictures were presented in the first phase of the experiment. The subsequent question was more specific and asked whether they noticed anything unusual about the presentation of words and pictures of people of different genders. Finally, I asked whether they knew the purpose of the study, and if yes, I asked for a brief summary. The tasks were programmed using DirectRT (Jarvis, 2012). The experiment took 20 minutes.

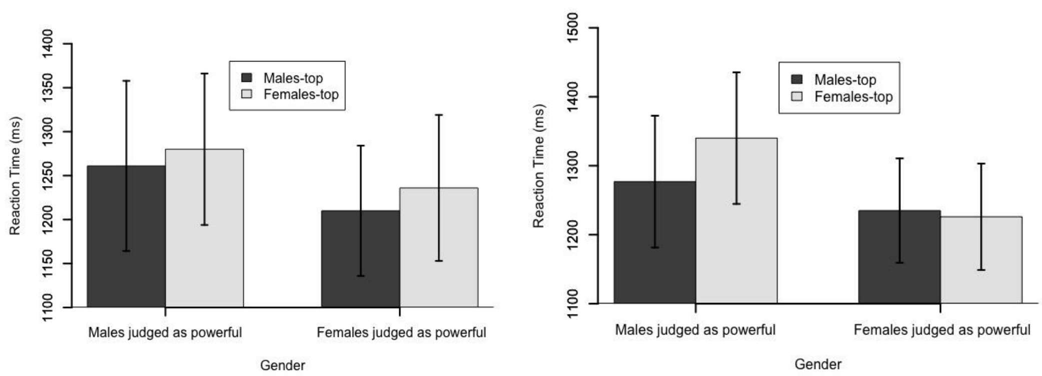
Results and Discussion

Proportions of choices. First, I analysed the proportions of participants' choices of males and females as powerful. As in Study 3, where I also paired gendered names with high-status professions, in both conditions (priming vs. control) participants were equally likely to select both males (52%) and females (48%) as powerful, $\chi^2(1, N = 43) = .08$, $p = .78$, $d = .11$; odds ratio $_{\text{choosing powerful in the primed versus control condition}}$ = .97 (95%CI[.80,1.18]).

Responses latencies. To analyse the response latencies, I estimated a linear mixed model. The fixed effects of prime, $F(1, 1599) = .02$, $p = .89$, $dz = .01$, trial type, *F*(1, 1600) = 1.46, *p* = .23, *dz* = .04, and gender choice, *F*(1,1599) = 1.61, *p* = .20, *dz* = .03 were not significant. The same was true for two-way interactions (prime x trial type; prime x gender choice; trial type x gender choice), *ps* = .31. Also, contrary to my hypothesis, the three-way interaction, prime by trial type by gender, was not significant, *F*(1, 1599) = .03, *p* = .86 (see Figure 19). It is possible that as a result of the high drop-out rate and also a small sample size, the present study is likely to be underpowered. A replication of the present study would be essential in order to deal with high variability found in participants' responses (see standard errors in Figure 19).

<u>Figure 19.</u> Response latencies as a function of the interaction among gender choice, trial type, and task for the control and prime condition. The error bars show standard errors (+/-1 SE).

(a) Control **(b)** Prime



Summary of results

In Study 11, I found that spatial associations between top and bottom positions were not moderated by stereotypic associations, as both males and females (when categorised as powerful) were associated with top and bottom positions to the same extent. Participants also selected more stereotype-consistent choices (powerful-males) in the long condition when they had more processing time available (Fiske & Taylor, 2010). In the short condition, however, the verticality cues were not associated with participants' responses. These findings are inconsistent with my initial studies (1 - 7), which suggested that only males as powerful were simulated at the top. At the same time, as mentioned earlier in Study 11, I measured only *associations* in terms of the frequency of participants' choices, which represent a different measure to reaction latency or pupil dilation. The data based on associations cannot reveal the effects due to processing effort. Therefore, a priming technique that would directly affect stereotype accessibility could be more useful in determining the role of stereotyping in spatial simulations precisely.

Nevertheless, the priming technique used to experimentally reduce stereotype-consistent thinking was not effective in Study 12. The results from Study 12 did not support any of the stated hypotheses. Also, I did not replicate my previous findings in the control condition. One possibility is that my priming method was not sufficient to stimulate formation of new associations between power and females and therefore the spatial effect was not stronger for female names than for male names in the priming condition. However, another probable explanation of my findings is that my sample size was not large enough. A replication of the above study with a larger sample size as well as a longer priming task (e.g., the same task that I used, but repeated twice) could potentially provide more reliable findings.

So far, I explored the context dependency of spatial simulations on task conditions and stereotype activation. In the next chapter, I aimed at investigating the consequences of spatial correlates of abstract thinking for gender perceptions.

Chapter 6: The impact of spatial vertical presentation on gender perceptions

Overview

In the final empirical chapter, I present two studies that investigated the consequences of vertical presentation on perceptions of males and females in terms of physical size and stereotypic traits. In Study 13, I explored whether the vertical presentation of stereotypically powerful persons at the top might bias a subsequent perception regarding size. Participants were primed with males-top trials (male names presented at the top and female names at the bottom), females-top trials (female names at the top and male names at the bottom), or no priming. Subsequently, participants were asked to estimate font sizes of target words which represented male and female names. I found that participants tended to overestimate the male names over female names regardless of condition. Moreover, they were more likely to overestimate both male and female target names after being primed with males-top versus (a) females-top (although this association was marginal) and (b) control trials. In Study 14, I explored whether the spatial simulation of stereotypically powerful persons (males) at the top might predict biased or stereotypically-consistent judgements about such persons. In this study, participants completed a standard spatial task (adapted form Study 1), in which they were asked to quickly select a powerful or powerless person on the screen on the basis of vertically presented pictures of male and female faces. Subsequently, participants rated the pictures presented in the spatial task on a range of stereotypical traits. The results indicated that participants simulated both males and females as powerful at the top versus bottom in the spatial task. These simulations were independent of participants' ratings of the faces traits. However, the higher the likelihood to select female faces as powerful at the top, the lower the femininity judgements of those faces, whilst a higher tendency to pick male faces as powerless at the bottom was associated with less competence and masculinity attributed to those faces. Overall, the studies indicate that verticality is associated with biased perception of size and stereotypic judgements.

Study 13: Space and size in the power-gender representation

My previous studies demonstrated that gender can be mapped onto the vertical dimension via another related concept, e.g., power. At the same time, the embodiment literature suggests that gender is likely to be directly mapped onto dimensions associated with bodily experiences. For example, Slepian et al. (2011) found that the proprioceptive experience of toughness is associated with males, whereas the proprioceptive experience of tenderness is associated with females. Such associations were present during an online thinking about gender when participants were squeezing a hard (symbolising toughness) or soft ball (symbolising tenderness). The findings by Slepian et al. (2011) indicate that bodily experiences might bias abstract judgments about gender. Similarly, a cognitive experience of being powerful or powerless was found to be related to judgments about other people's physical sizes. In one piece of research, Yap, Mason, and Ames (2013) manipulated participants to feel either powerful or powerless and asked them to judge the physical size of a target person with whom they interacted. The results indicated that participants who felt powerful tended to underestimate the physical size of the target person, whilst those who felt powerless were more likely to overestimate the size of that person. That is, perceptions of physical sizes were biased by abstract cognitive correlates of power.

Most studies of congruency effects focus on mappings of a single abstract concept onto a specific dimension of space (e.g., Boroditsky Ramscar, 2002; Meier & Robinson, 2004; Schubert, 2005). Such studies indicate that different abstract concepts, such as valence, time, or power, use the same spatial mappings. However, not many studies have investigated whether multiple mental metaphors can be activated at the same time and whether they interact with each other to produce a coherent mental representation. Spatola et al. (2018) examined these issues in the context of metaphors in which negativity is associated with past events, whilst past events are also associated with the left side on the horizontal dimension. In their studies, participants made either temporal (past vs. future) or valence (positive vs. negative) judgements about verbs presented in the past or future forms. Laterality of responding was also manipulated, such that participants made their responses with keyboard keys positioned either on the left or right. The authors measured how accurate and fast participants were when making such judgements in order to measure congruency effects between the spatial dimensions and the abstract concepts (see also Chapter 1). They found that participants showed a facilitation while responding to past-negative/future-positive verbs, past-left/future-right verbs, and left-negative/right-positive verbs. Although these three mappings (valence-time, time-space, and space-valence) were activated simultaneously such representations were independent of each other. These findings suggest that people process multiple mental metaphors globally. Within a mental model, people activate attributes associated with each of the metaphors (left-negative, past-left, past-negative) and keep them active throughout the task, creating a coherent mental representation. Under such circumstances, the information about mappings is available in advance. However, it is important to note that in this research multiple metaphors needed to be activated simultaneously, due to the nature of the task. Activating metaphors in a sequential manner might yield different results, because one mental representation would be formed in advance of the other. I addressed this issue in the current study.

Integrating the findings by Slepian et al. (2011) and Yap et al. (2013) on the association between gender, embodied experience, and physical size with the studies on spatial mental representation of abstract concepts (e.g., Meier & Robinson, 2004;

Schubert, 2005; von Hecker et al., 2013), and the lack of interaction among simultaneously activated metaphors (Spatola et al., 2018), I investigated whether two aspects of space, i.e., vertical location and size, might correlate to bias people's perception of a social category. Given that power is associated with both spatial location and size, it is possible that these two aspects of space interact in mental representations of power – perceptions of individuals in higher vertical locations might also lead individuals to perceive other stimuli as physically bigger. Given the findings of Spatola et al. (2018), such an interaction between metaphors should only be possible when the mental representation of one metaphor (power-space) is activated before the other one (power-size).

My research questions were motivated by two overarching issues. First, if location and size are tied to the concept of power (Lakoff & Johnson, 1999; Schubert, 2005; von Hecker et al., 2013), then when power is attributed to social categories like gender, the mental representation of gender would also involve associations with both location and size. Further, as socially constructed beliefs about power indicate that males are usually considered as socially more powerful than females within Western societies, males should be more readily associated with power than females (Rudman & Kilianski, 2000; Rudman et al., 2012). This was demonstrated in my initial studies (1 - 7), which indicated that concrete spatial thinking is pronounced only when people think in stereotype-consistent ways. Second, if bigger sizes are associated with greater power, then seeing males as more powerful should be also associated with making members of the same category appear bigger than they are in reality. Finally, if spatial location and size are correlated in mental representations of power-gender associations, male names (rather than female) presented in upper locations should affect size perceptions such that subsequent stimuli should appear to be bigger than they are in reality.

Therefore, in Study 13, I investigated whether the vertical associations in power-gender representations would also be associated with size. This was tested by having participants complete a computer task where they estimated the font sizes of gendered names. Before each font judgement, some participants were primed with vertically positioned male and female names. In the males-top condition, the male name appeared at the top, whereas in the females-top condition, the female name appeared at the top. Participants were informed that those names represented managers and subordinates (to activate power-relevant associations). I also included a control condition that did not involve any priming (see Figure 19). As gender has power implications, I predicted that male names should be overestimated to a higher extent than female names. Also, I tested whether male names would be overestimated more in the males-top condition compared to the females-top and control conditions; the same was tested for female names in the females-top condition.

Method

Participants

I recruited 87 Cardiff University undergraduate students (67 females; mean age = 20.33) and randomly assigned them to one of the three between-subjects conditions: males-top condition, females-top, and control. The sample size was derived from von Hecker et al. (2013, 2016) for Cohen's d = .25-.35. They received course credit for their participation.

Materials

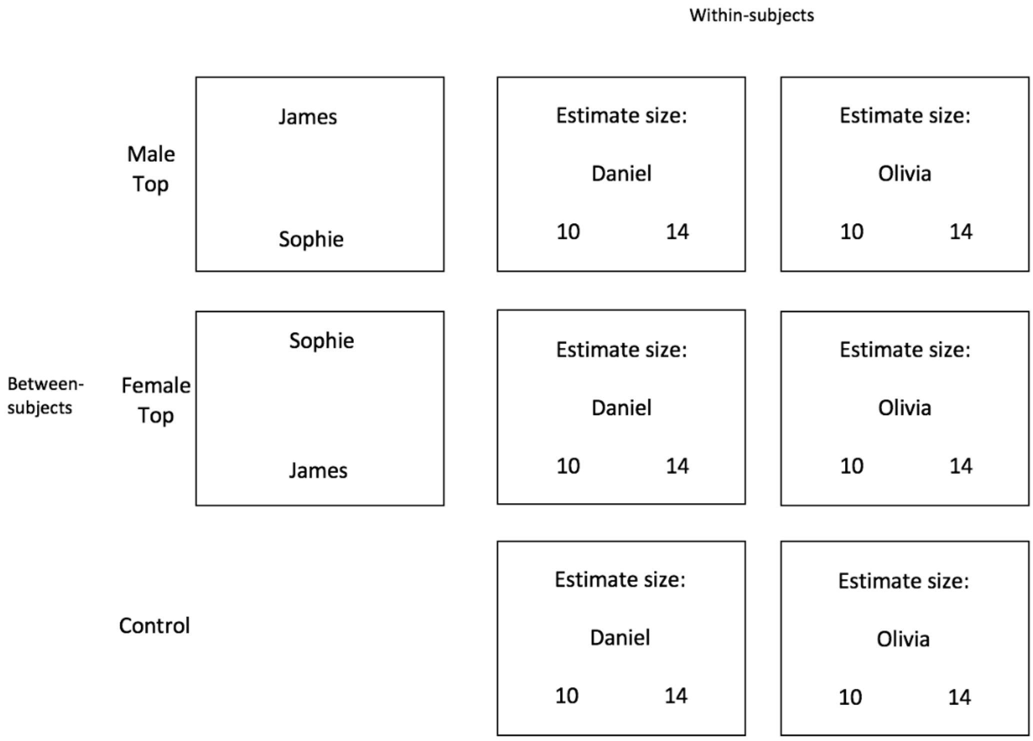
Participants completed the estimation task on a computer screen using DirectRT (Jarvis, 2012). Within all conditions, I presented a practice phase, where participants estimated the font size of neutral words. In each of four blocks participants estimated the font size of six neutral words (e.g., tree, mushroom, fruit). Next, they completed the main task, in which they estimated the font size of gendered names across six blocks. These consisted of six popular male British names (Matthew, Ryan, Daniel, Henry, Nathan, Gabriel) and six popular female British names (Zoey, Joanna, Olivia, Natalie, Victoria, Ann). There were 12 trials per block (72 trials in total). Each stimulus within the practice and experimental phase was presented in six different font sizes (12, 14, 16, 18, 20, and 22 points). Participants were given two font estimation choices. The first choice referred to a font size that was two points smaller than the actual size of the presented word, whilst the second choice was two points larger than the actual font size of the word (for the word displayed in font size 12, participants could choose font 10 or 14). Therefore, participants could not provide the correct answer, but were forced to either underestimate or overestimate the word's font size.

In the two experimental conditions (males-top and females-top), before the estimation task, participants were primed with vertically arranged gendered names. These names were also selected from a list of the most popular British names (the names were selected from the same sources as the names presented in Studies 1 - 7). Subsequently, I paired a female name with a male name and always presented the same pairs throughout the experiment (see Appendix 3: Studies 1 and 3). Each trial of the main estimation task began with a display of a male-female pair, where either the male name appeared at the top and female name at the bottom (e.g., Oliver-top; Emily-bottom; males-top condition), or the female name appeared at the top and the male name at the bottom (e.g., Emily-top; Oliver-bottom; females-top condition). To control for participants' awareness of the link between the vertical presentation of the names and the subsequent estimation task, I informed them that I was interested in their memory performance and they should try to memorise the vertically presented names. At the end of the study, they were asked to list all the memorised names. The instructions and the answers that participants could select were always displayed in the same font size (15 points). All the stimuli and instructions were presented in white letters on a black screen.

Design

I used a mixed-design. I manipulated the vertical position of the gendered name pairs between-participants; there was also a no-prime control condition (see Figure 20). The target names were manipulated within-participants and were male or female. Across all conditions, I measured the proportions of participants' underestimated and overestimated font sizes. The trials within each block were randomised.

Figure 20. The procedure of a single trial in the estimation task.



Procedure

Participants within all conditions completed the practice phase followed by the main task. In the practice phase, each trial began with information that the trial was about to start (for 2000ms) which was followed by a blank screen (displayed for

1000ms). Afterwards, the target word was displayed. Below each word, I presented two possible font size estimates (e.g., font size 10 and 14) that participants could select, one on the left, and the other on the right. Participants pressed the arrow pointing left to pick the number on the left and the arrow pointing right to pick the answer on the right. When they had made their response, the next trial began. Each trial in the main task started in the same way as in the practice phase, however, before the estimation task began, participants within the experimental conditions were informed that their task would involve memorising pairs of names displayed on the computer screen and that these pairs would be presented before the estimation task within each trial. To activate the context of thinking about power, I informed participants that the pairs of names would represent people who were managers and subordinates. First, in each trial, participants were asked to fixate on mark X in the middle of the screen for 1000ms. Second, a pair of gendered names (with one name displayed at the top and the other at the bottom) appeared for 3000ms. Then the estimation task followed and participants had unlimited time to estimate the size of the displayed target name. In the final stage, participants were asked to write down all the memorised gendered pairs on a sheet of paper. The experiment lasted approximately 25 minutes.

Results and discussion

I analysed the proportions of participants' font size estimations (i.e., the frequency of overestimated and underestimated sizes) across three conditions using loglinear analysis.[31] I subjected three variables to the analysis: Condition (three levels: males-top, females-top, and control), gender name (male or female), and estimation (overestimated and underestimated of size). First, I found a main effect of estimation, $\chi^2(1, N = 87) = 102.68$, $p < .001$, $d > 2.0$. The analysis revealed that participants were

[31] I conducted the same analysis excluding male participants in the sample. I obtained the same results as in the analysis that included both female and male participants.

more likely to underestimate (56.0%) rather than overestimate (44.0%) all names (odds $_{\text{underestimation versus overestimation}}$ = 1.29, SE $_{\text{proportion}}$ = .06).

Next, as hypothesised, I found a significant main effect of names on estimation, $\chi^2(1, N = 87) = 14.07$, $p < .001$, $d = .88$, such that participants were more likely to overestimate male names (45.9%) compared to female names (41.3%; odds ratio $_{\text{males versus females overestimation}}$ = 1.21, 95%CI[1.10,1.34]). Further, as expected, I found that the main effect of condition on estimation was significant, $\chi^2(1, N = 87) = 16.24$, $p < .001$, $d = .96$. Although participants were more likely to underestimate the font size, they did this more frequently in the control condition as compared to the experimental conditions. Follow-up tests revealed that the lowest percentage of underestimated names was observed in the males-top condition (46.6%) as opposed to the control condition (40.5%), $\chi^2(1, N = 58) = 16.13$, $p < .001$, $d = 1.24$ (odds ratio $_{\text{males-top versus control overestimation}}$ = 1.29, 95%CI[1.14,1.45]). Likewise, participants underestimated the names less in the females-top condition (43.9%) than in the control condition, $\chi^2(1, N = 59) = 5.20$, $p < .02$, $d = .63$ (odds ratio $_{\text{females-top versus control overestimation}}$ = 1.15 (95%CI[1.02,1.30]). There was also a marginal tendency to underestimate the names less frequently in the males-top condition compared to the females-top condition, $\chi^2(1, N = 57) = 3.04$, $p = .08$, $d = .47$ (odds ratio $_{\text{males- versus females-top overestimation}}$ = 1.12, 95%CI[.99,1.26]). This pattern of results was true for both types of analysis: collapsing across gender names and testing for their association with condition as well as when each gender name was analysed separately. Finally, the three-way interaction among gender name, condition, and estimation was not significant, $\chi^2(1, N = 87) = .20$, $p = .91$, $d = .10$.

To summarise, in the present study, I showed that spatial implications of power have consequences for the perception of physical sizes, such that a presentation of stereotypically powerful individuals at the top increases the perceived size of subsequently presented male and female names. Given that the vertical positioning biases people's perception of another aspect of space (i.e., size), in the subsequent study I explored whether such spatial arrangement can be also involved in biased *abstract* and *social* perception of the presented persons.

Study 14: Consequences of spatial simulations for other stereotypic judgements

I designed Study 14 with two aims. First, I wanted to investigate whether the spatial simulation of gender might impact participants' stereotypic judgements about the simulated individuals. According to the Stereotype Content Model (SCM, Fiske et al., 2002), stereotypes of social groups are best described using two dimensions: warmth and competence. This model suggests that socially undesirable consequences of stereotyping, such as justification of prejudice, might stem from both negative and ostensibly positive views of social groups. For example, females or Black people are perceived as being incompetent, which is a negative trait that could lead to discrimination of such individuals at the workplace (Haddock & Zanna, 1994; Katz & Hass, 1986). At the same time, these groups are also considered to be high on warmth and therefore they are treated in a paternalistic way (deserving pity and being in need for protection or help; Glick & Fiske, 1996). The SCM suggests that such mixed stereotypes can be predicted from the social structure variables associated with social status and competition. Perceived competence is correlated with high social status while low warmth is associated with high social competition. As a result, groups that hold a high social status and are powerful are also perceived as being competent. In turn, subordinate groups that are less likely to compete are considered to be warm. This serves a social function of maintaining the social structure – as long as the subordinate groups are treated in paternalistic way and are considered as "nice," they would be less likely to attempt to change their status and compete with powerful groups (Glick &

Fiske, 2001). Similar processes can be observed in terms of sexism. Glick and Fiske (1996) provided evidence for a mixed nature of prejudice against women. In general, in their study, women were perceived as incompetent in a professional context. At the same time, participants considered them to have many positive traits, which conformed to female traditional gender roles, that is, being nice and warm, but dependent on men. Again, such mixed stereotype content justifies and maintains social inequality. Overall, it appears that stereotype content is predicted by social competition and status. Therefore, in the present study, I explored whether spatial location interacting with perceived power differences between male and female targets might bias participants' judgements about the target's perceived warmth and competence.

As warmth and competence are two basic dimensions that capture stereotype content, and groups stereotyped as powerful (e.g., males) are represented at the top of the vertical dimension (see Chapters 2 - 3 in this thesis), I predicted that simulations of males as powerful at the top (i.e., faster response to males at the top versus bottom) would be associated with judgements of more competence and less warmth of those males in contrast to males simulated at the bottom. This is because the stereotype associated with each gender should be reinforced or weakened depending upon the spatial position. The upper vertical position would reinforce the stereotype of being competent when stereotype-fit between males and power would be present. In contrast, the lower position would reinforce the stereotype of being warm, when stereotype-fit between females and lack of power would be activated. That is, the simulated powerful or powerless females at the bottom versus top would be associated with higher warmth and lower competence judgements.

Second, in my previous studies, I used gendered names or gendered names paired with professions as stimuli to test whether people mentally simulate males/females as powerful/powerless on the vertical dimension. Also, in Study 12, which used names as stimuli, I did not find any evidence for reduced spatial simulation effects when participants were primed with females as powerful and males as powerless relative to the condition where no priming took place. It is likely that stimuli demonstrating a more vivid or salient representation of gender (e.g., pictures) might be more easily mapped onto people's internal spatial representations (Nisbett & Ross, 1980). Subsequently, if a presentation of more vivid stimuli results in more pronounced spatial representations, then it would be also possible that such representations would have more impact on people's cognition and affect their subsequent judgements to a higher extent. Therefore, in Study 14, I used pictures of neutral male and female faces. If presenting pictures is associated with more pronounced spatial simulations than names, I also predicted a simulation of females as powerful at the top versus bottom, but I expected this effect to be weaker than the effect for males as powerful at the top. I also expected to find simulations of females as powerless at the bottom as opposed to the top, which should be more pronounced for females judged as powerless as opposed to males judged as powerless. I hypothesised that faster reaction time to choices of females as powerless when their names appeared at the bottom as opposed to the top would be associated with perceptions of lower competence and higher warmth of the simulated females. Specifically, I expected that stereotypic thinking that involves spatial simulations would enhance stereotypic associations among power, gender, warmth and competence. Additionally, I also explored whether such simulations would be associated with participants' liking of the simulated individuals and their perceived masculinity or femininity.

Method

Participants

I recruited 71 Cardiff University undergraduate students (mean age = 19.60; 50 females, 5 males, one non-binary, and 15 did not report their gender). The sample size was derived from von Hecker et al.'s study (2016) for Cohen's d = .25-.35. Four participants misunderstood the instructions, whilst the data of four other participants were missing, so the final sample consisted of 63 participants. They were randomly allocated to either the powerful or powerless condition and received course credits for their participation.

Materials

Spatial task. I selected pictures of female and male faces from Radboud Faces Database (Langner et al., 2010). To avoid any potential effects of facial expressions, attractiveness, or age, I chose pictures that represented neutral faces of individuals of roughly the same age (range: 23-28) and who were matched on attractiveness (rated as neutral). After pre-selecting the pictures, I randomly chose five pictures of females and five pictures of males. Subsequently, I matched the pictures of females with picture of males in any possible combination within four sets of pairings (as it was done for gendered names and professions, see Study 3). The size of each picture was 400 x 280 pixels. The pictures were presented against a white background.

Judgement task. The same pictures, as presented in the spatial task, were used in the judgement task. Participants were told to assess each individual in terms of warmth, competence, likability, and femininity or masculinity, on a scale from 1 (not warm/competent/likable/feminine or masculine at all) to 7 (very warm/competent/likable/feminine or masculine). The presentation order of pictures and questions was randomised, however, participants always answered four questions in a row about the same picture. The questions were presented in black letters against a white background (font 20 points) with presentation software MediaLab (Jarvis, 2012).

Participants selected their responses by indicating a number on a scale with a mouse. They saw one question at a time. They had unlimited time to provide a response.

Design

In the spatial task, I randomly allocated participants to one set of four combinations of matched pictures (one was always representing a female face, whilst the other one a male face) and also to the powerful of powerless condition. I presented four blocks of trials that included 10 trials each. In each block, five trials presented five pictures of males at the top, whilst their paired pictures of females were presented at the bottom. On the other five trials, male pictures were at the bottom, whilst female ones at the top. That is, trial type was manipulated (males-top versus females-top). The trials were randomised within each block. I measured how often participants indicated males/females as powerful/powerless in each vertical position and also their reaction times to those choices (gender choice was the quasi-experimental variable like in my previous studies presented in Chapter 2). Before each trial, participants saw a fixation mark in the middle of the screen. The timing was the same as in the previous studies (see Figure 1).

Procedure

First, participants were asked to either quickly find the powerful or powerless person on the screen. The definition of a socially powerful/powerless person was provided (as it was done in the previous studies, e.g., in Chapter 2). Participants were encouraged to imagine that one of the presented persons on the screen was powerful or powerless and think strongly about the power differences between them. Subsequently, they completed the spatial task and after that answered questions about the pictures presented in the spatial task. The experiment took 20 minutes.

Results

Spatial task. Proportions of choices. Similarly to the previous studies, I found that participants were more likely to pick males as powerful (55.4%) versus females as powerful (44.6%), $\chi^2(1, N = 63) = 18.35$, $p < .001$, $d = 1.28$, odds ratio $_{\text{choosing males over females as powerful}}$ = 1.42 (95%CI[1.21, 1.66]).

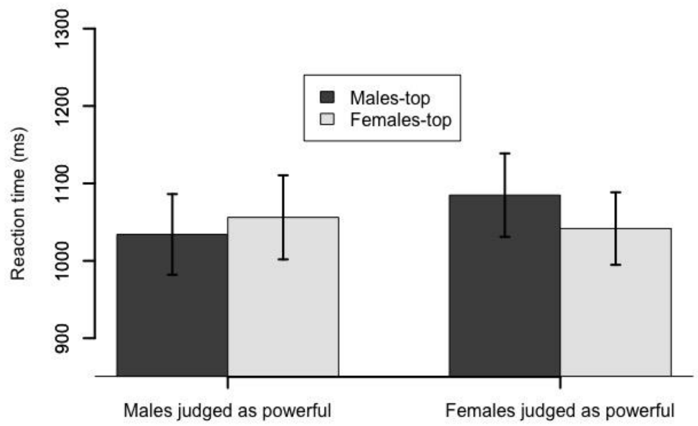
Spatial task. Response latencies. As in previous studies, I estimated a linear mixed model to analyse participants' reaction times to their choices of males and females as powerful or powerless. I found that the fixed effects of gender choice, $F(1, 61) = .02$, $p = .89$, $dz = .01$, trial type, $F(1, 2337) = 1.11$, $p = .29$, $dz = .07$, as well as task, $F(1, 61) = 2.54$, $p = .12$, $dz = .07$, were not significant. The same was true for the two-way interactions (gender choice x trial type; gender choice x task; trial type x task), *ps* > .24.

Further, as predicted, the three-way interaction (gender choice, trial type, and task was significant, $F(1, 2309) = 8.41$, $p < .01$ (see Figure 21). However, post-hoc comparisons indicated that participants were equally fast to categorise males as powerful, ($M = 1034$, $SE = 63.75$, 95%CI[915, 1170]), and females as powerful at the top, (M = 1042ms, $SE = 66.20$, 95%CI[898, 1137]),]), $t(2319) = -1.96$, $p < .05$, $dz = .08$, in contrast to the findings in Studies 1 – 6. I also did not find any spatial effects in terms of females as powerless at the bottom (M = 1168ms, $SE = 60.29$, 95%CI[1083, 1323]) versus the top, (M = 1170ms, $SE = 61.32$, 95%CI[1082, 1327]), $t(2342) = -.06$, $p = .95$, $dz = .03$.

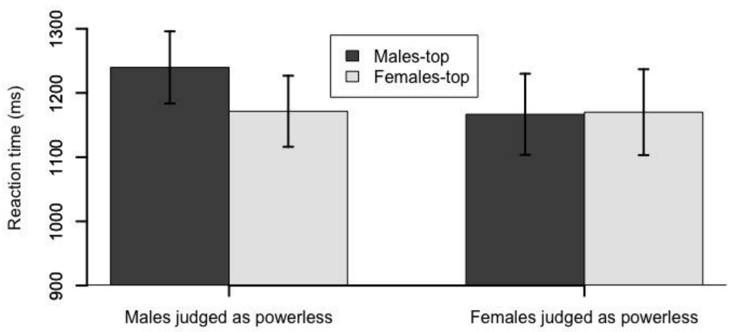
Subsequently, I found that participants were faster at detecting males as powerful at the top ($M = 1034$, $SE = 63.75$, 95%CI[915, 1170]) versus bottom, (M = 1056ms, $SE = 64.35$, 95%CI[961, 1218]), $t(2319) = -1.96$, $p < .05$, $dz = .28$. The same was true for females as powerful, they were chosen significantly faster at the top, (M = 1042ms, $SE = 66.20$, 95%CI[898, 1137]), versus bottom, ($M = 1085$, $SE = 71.29$, 95%CI[956, 1199]), $t(2319) = 2.19$, $p < .03$, $dz = .31$. Other comparisons were not significant, *ps* > .20.

Figure 21. Response latencies as a function of the interaction among gender choice, trial type, and task. The error bars show standard errors (+/-1 SE).

(a) Powerful condition



(b) Powerless condition



Judgement task. Due to programme failure, the questionnaire data of 13 participants were missing. Therefore, I analysed the data from 50 participants. To test whether spatial simulations predicted participants' judgments about perceived warmth and competence of the simulated individuals, first, I computed spatial simulation-facilitation scores. As I did not find any spatial effects for choices of powerless individuals at the bottom versus top, as it was the case in the previous studies, I computed the facilitation scores for males (FTM) and females (FTF) chosen as powerful when their name appeared the top. To do that, I calculated mean reaction times for each participant to the choices of males as powerful and females as powerful at the top and subtracted these means from the mean reaction times to the choices of males as powerful and females as powerful at the bottom, respectively. That is, higher facilitation scores indicated faster responses to males/females as powerful at the top versus bottom. Subsequently, as participants rated each picture of female and male face in the judgement task, I calculated the mean rating for each questionnaire item (warmth, competence, likability, and femininity/masculinity) for all female faces and separately for all male faces giving an overall rating of each item for each gender (see Appendix 9: Table 4 for correlations between all items).

I found that FTM did not predict warmth, $\beta = -.34$, $t(26) = -1.57$, $p = .13$, competence, $\beta = -.10$, $t(26) = -.49$, $p = .63$, and masculinity $\beta = -.03$, $t(26) = .14$, $p = .89$, judgements of males. However, the facilitation score significantly predicted likability ratings, such that the faster the response to males as powerful when their name appeared at the top, the less likable the male faces appeared to participants, $\beta = -.44$, $t(47) = -3.56$, $p < .01$. The facilitation to females as powerful when their name appeared at the top (FTF) did not predict any of the judgements: warmth, $\beta = -.11$, $t(47) = -.11$, $p = .47$, competence, $\beta = -.44$, $t(26) = -2.47$, $p = .02$, femininity, $\beta = -.17$, $t(45) = -1.12$, $p = .27$, and likability, $\beta = -.17$, $t(47) = -1.19$, $p = .24$.

Exploratory analysis. Although the simulation scores did not seem to predict participants' judgements of the simulated individuals, I also explored whether the proportion of choices (males-top/bottom; females-top/bottom) in two conditions (powerful versus powerless) would be related to such responses. As mentioned before, in Study 11, I found that the mental representation of power involved spatial features whenever participants indicated males or females as powerful. Therefore, for each participant, I calculated how likely they were to pick each possibility (males at the top/bottom and females at the top/bottom) out of four possible choices within the powerful/powerless condition. Therefore, if they were equally likely to pick each one of them, the likelihood of their choice would be 25% for each category.

First, in the powerful condition, I found that the higher likelihood of choosing females as powerful at the top was negatively associated with femininity judgements, specifically, participants considered them as less feminine, $r(28) = -.45$, $p < .02$. This negative correlation was specific to the top position, as no such relationship was found between the femininity judgements and the likelihood of categorising females as powerful at the bottom, $r(26) = -.22$, $p < .29$.

Second, in the powerless condition, I found that higher likelihood of choosing females as powerless at the bottom was positively associated with warmth judgements, $r(22) = .54$, $p < .01$. This was not specific to the position, as the same applied to females as powerless chosen at the top, $r(21) = .49$, $p < .03$.

Third, in the case of male targets selected as powerless, there was a negative association between the likelihood of choosing them at the bottom and their perceived competence, $r(22) = -.45$, $p < .04$, and masculinity, $r(22) = -51$, $p < .02$. Specifically, the higher the tendency to select males as powerless at the bottom, the lower their rated competence and masculinity. These findings were also specific to the bottom position, as other correlations were not significant, $ps > .14$.

Summary of results

In the present chapter, I aimed at investigating whether vertical presentation of gendered names would have an impact on people's stereotypic perceptions of gender. In Study 13, I supported my hypothesis by indicating that in general males are considered to be bigger than they are in reality in comparison to females. Importantly, I found that males presented at the top primed less size underestimation of both male and female names than when no priming occurred. Also, males-top primed marginally higher overestimations of both types of names than females-top priming. Overall, I demonstrate that presenting a stereotypically-consistent person at the top biases people's perception of physical size, such that top position correlates with perceptions of larger size.

As Study 13 demonstrated that verticality biases concrete perceptions of gendered individuals in terms of their physical size, I designed Study 14 to see whether verticality can also bias abstract stereotypic judgements. First, I explored whether spatial simulations might bias participants' judgements about the simulated individuals in terms of two basic dimensions of stereotypes, warmth and competence, as well as likability and femininity for female individuals and masculinity for male ones. I also tested whether presenting pictures instead of gendered names would be associated with stronger spatial simulations due to enhanced vividness of pictures. My predictions were partly supported. As hypothesised, I detected spatial simulations of male and female faces as powerful at the top versus bottom. However, participants were not faster at categorising male faces as powerful when they appeared at the top, as opposed to categorising female faces as powerful in the same location. Again, no spatial associations emerged when participants categorised individuals as powerless when their faces appeared at the top and at the bottom.

In terms of participants' stereotypic judgements, overall, the simulation scores to males and females as powerless at the top did not predict competence and warmth judgements. The same was true for femininity and masculinity. However, participants who were quicker to detect the males as powerful at the top versus bottom, disliked those males more. This may imply that participants might not accept the concrete spatial association of males as powerful at the top. This could be attributed to a sample-specific effect – mainly female participants contributed to these results suggesting that they might have disliked considering males as powerful. Subsequently, in my exploratory analysis, I found that participants' ratings in the judgement task were associated with the likelihood of their choices of males/females in each location within the spatial task. That is, the higher likelihood of indicating females as powerful at the top was associated with their lower femininity ratings. Also, indicating them as powerless at the bottom or the top was associated with their higher perceived warmth. Finally, males as powerless indicated at the bottom, but not at the top, were judged as less competent and masculine. In general, higher proportions of choosing females as powerless were associated with higher scores of warmth, which is consistent with SCM (Fiske et al., 2002).

Overall, when people perceive a power difference between genders, spatial location indeed interacts with gender stereotypes. The interaction involves reinforcing stereotype-consistent beliefs about warmth and competence. However, it is essential to note that the association between spatial locations and stereotype-consistent thinking is only observed between the *choices* that participants make regarding the powerful/powerless individual in a specific location, but not the *simulation* of those individuals. That is, the speed with which people process stereotype-consistent person in a specific location does not predict subsequent stereotypic judgements. It is rather the outcome of the processing, i.e., the judgement, that people make about power and location that correlates with warmth and competence judgements.

Chapter 7: General Discussion

Overview

In the last chapter of my thesis, I integrate and discuss the findings obtained in 14 experiments. First, I outline the main aims of the present thesis that involved investigating (a) the presence of spatial simulations in the mental representation of gender stereotypes in the context of social power; (b) whether spatial simulations depend on the power-context and stereotype accessibility, and (c) consequences of spatial simulations for gender perceptions. Then, I summarise my results and discuss the findings of each set of studies presented in empirical chapters. I subsequently present limitations and alternative theoretical perspectives. Finally, in the concluding section, I integrate my main findings suggesting that sensorimotor spatial simulations are present in the mental representation of gender via perceptions of power.

The main aim of my thesis was to investigate whether abstract reasoning about social categories involved sensorimotor spatial simulations, moderated by stereotypic thinking. Specifically, I wanted to examine the precise representational features associated with gender stereotypes in relation to two theoretical perspectives on cognitive processing - grounded cognition and conceptual blending. Overall, across 14 experiments, I show that the mental representation of gender-power stereotypes is associated with sensorimotor simulations only when people select stereotype-consistent stimuli (males as powerful). Participants spatially simulated male names opposed to female names when the names appeared at the top versus the bottom only when they categorised them as powerful. When stereotype-inconsistent categorisations were made (females-powerful), no such effects occurred, except for when participants were making judgements about pictures of male and female faces rather than names (Study 14). The effects were not present in the case of thinking about powerless individuals. These findings were based on reaction-time methodology, but were further replicated using pupillometry. This method revealed that making judgements about metaphorically incongruent as compared to congruent concepts (powerful-males-bottom; powerful-bottom) involved increased cognitive conflict and effort as indexed by increased pupil size.

Subsequently, I found that spatial congruency effects are dependent on task features. Spatial simulations only occurred in semantically meaningful contexts, and gender was only spatially simulated when males and females were compared against each other. Also, gender per se, when categorised without the power context, was not simulated vertically, suggesting that verticality cues are not enough to elicit power-related spatial simulations (Study 10). Interestingly, however, I found that overall spatial cues were in fact associated with the *type* of response participants made. This happened only in Study 9, where participants categorised just gendered names (i.e., not paired with professions as in Study 8). This was because participants tended to select all names (both male and female) as powerful more often at the top versus bottom and vice versa when choosing them as powerless. This was most pronounced in the case of categorising *males*, but not females. These choices were not associated with faster reaction times.

Furthermore, I found that both the spatial simulations (associated with the time of responding) and the type of power categorisations people make (associated with the outcome of processing) do not depend on the level of stereotype activation. Priming participants with powerful-females/powerless-males did not involve spatial simulations of females at the top or reduced simulations of males at the top. That said, these results were based on an underpowered study that also failed to replicate the general powerful-males-top effects. Therefore, these results should be treated with caution. Moreover, restricting decision-making time was associated with fewer stereotype-consistent choices. However, such restriction did not moderate space-power associations in the type of categorisations. Taken together, these findings suggest that the *type* of decisions people make about power involves stable space power associations. It is more likely that stereotype activation impacts *spatial simulations,* which are associated with the duration of processing. This is because in Studies 1 – 7, simulation was observed when a stereotype-consistent judgement was made (i.e., males as powerful). Such findings are in line with the conceptual blending framework as they suggest that combining a concept that has sensorimotor features (i.e., power and verticality associations) with a conceptually consistent stimulus (gender linked with power via stereotypes), should lead to a spatial simulation of that stimulus (powerful-males-top bias).

Finally, having knowledge about the reliability of spatial simulations in the context of gender-power associations, I investigated whether such simulations might influence perceptions of simulated individuals. I observed that presenting a stereotypically fitting person at the top (i.e., powerful-males) was associated with less size underestimations of subsequently presented male and female names. The extent of this effect was the highest when people were primed with male names positioned at the top of the screen, as opposed to female names at the top or when no priming occurred. Next, categorising female faces as powerful when they appeared at the top was associated with perceiving these faces as less feminine. Similarly, when male names presented at the bottom were judged as powerless, the target was associated with lower perceived masculinity and competence. These findings show that the type of response people make involves sensorimotor spatial features that seems to reinforce stereotype-consistent judgements about represented individuals. Below, I provide a detailed discussion of the studies presented in each empirical chapter. The final conclusions of each empirical chapter and an overall interpretation of my findings is presented at the end of the chapter.

Stereotypic thinking involves spatial simulations - Chapter 2

The studies presented in Chapter 2 were designed to test whether cognitive blending between power and gender is associated with spatial simulations, given that power can be applied to gender via socially-constructed stereotypic beliefs. I explored whether the perception of gender in stereotypic ways would employ cognitive blending and therefore be associated with spatial simulations. Overall, I found that conceptual integration of power and gender is present when participants perceive stereotype-fit between gender and the individual judged as powerful (i.e., males versus females). I present evidence for spatial simulations when power level is judged in each trial by participants themselves, with respect to target concepts that are malleable in terms of power association (gendered names). Thus, power is applied (and spatially simulated) for gender in a malleable and situation-dependent fashion that depends on whether stereotypic thinking takes place or not in a given context, that is, for a given stimulus combination in my task.

My results indicate that participants were marginally faster to make decisions about the socially powerful or powerless person when the male name appeared on the top of the screen compared to when the female name appeared on the top of the screen. Further, there was a marginal effect of gender choice, as participants were marginally quicker to categorise male than female names. Also, male names were chosen significantly faster on males-top trials versus females-top trials. These findings indicate a general tendency to respond faster when being presented with a male name located above a female name. These tendencies were then moderated by task, as demonstrated by the three-way interaction. Consistent with expectations, I found that participants were significantly quicker at detecting males as powerful when their names appeared at the top as opposed to females as powerful at the top. Further, this effect was associated specifically with vertical location, as males were categorised as powerful significantly faster when they appeared at the top than bottom. The same pattern of results was not true for female names. They were categorised as powerful equally fast at top and bottom. No effects emerged in the case of selecting male and female names in the powerless condition.

Additional support for the involvement of stereotypic processes in spatial simulations was found in relation to the proportion of stereotype-consistent and stereotype-inconsistent choices. Participants were more likely to indicate stereotype-consistent choices (e.g., males as powerful; females as powerless) than stereotype-inconsistent ones. They were also faster at doing so – participants were faster to detect males as opposed females as powerful. However, the same was not observed for the powerless condition – participants were equally fast to categorise male and female names as powerless. Overall, these findings suggest that when people think about power in the context of specific social categories such as gender, they are more likely to apply the concept of power in stereotype-consistent ways, which is then associated with spatial simulations. However, this is pronounced only in the case of thinking about gender in terms of *power*, but not the lack thereof.

However, the tendency to make stereotype-consistent choices does not preclude the fact that sometimes participants also made stereotype-inconsistent choices. I argue that such choices would be affected by contextual processes such as vertical cues (top stimuli would receive an advantage over bottom stimuli) and participants' knowledge concerning power associated with individual gendered names. Wilder (1981) suggested that stereotyping is more likely to take place when people make judgements about entire groups rather than an individual. That is, individuation should reduce stereotypic thinking (see Ambady, Paik, Steele, Owne-Smith, & Mitchell, 2004, for empirical evidence). Hence, it is possible that perceiving specific gendered names might act as an individuation. Further, the literature indicates that the stereotype activation and application might depend on specific goals or the motivation to avoid prejudice (Kunda & Spencer, 2003). It is possible that some participants might have been motivated to avoid prejudice, which reduced the accessibility and application of a gender stereotype to their power judgement.

I did not find spatial simulations in the case of thinking about males as powerless, and females as powerful and powerless. This suggests that the simulation of a specific social category takes place only when participants think about it as being powerful, but not powerless. Although the pattern of participants' response latencies was consistent with my predictions (when the name at the bottom of the screen was selected as powerless, the response time was faster when it was a female name compared to a male name), the post hoc comparisons showed that the differences in RTs between responses as a function of verticality were not significant. These findings do not converge with Schubert's (2005) results, who found spatial simulations of powerless groups at the bottom. However, it is worth noting that Schubert's data were obtained in the context of highly-defined social groups in terms of power (*master* versus *servant*), whereas I used more malleable social categories, which might, per se, not be as clearly associated with high or low levels of power.

There are a number of plausible explanations for the lack of spatial simulations of females as powerless. First, extant research has demonstrated that spatial effects are usually more pronounced for linguistically unmarked concepts (e.g., powerful) than for marked ones (e.g., powerless). This is in line with the idea that unmarked concepts are processed faster than marked concepts due to familiarity, evaluative implications, and higher linguistic frequency (Hamilton & Deese, 1971; Meier & Robinson, 2004; Meier, Hauser, Robinson, Friesen, & Schjeldahl, 2007; Proctor & Cho, 2006; Schubert, 2005; von Hecker et al., 2016).[32] Second, it is worth noting that Schubert (2005) also found weaker effects in spatial simulations of powerless individuals relative to powerful individuals (see also Meier & Robinson, 2004; von Hecker et al. 2016).

In addition, research suggests that people are significantly slower when they make lexical decisions about negative words than positive words (see Estes & Adelman, 2008, for an overview). On the one hand, the effects found in my studies could be attributed partially to the valence, given that in general powerless groups were considered more negative. Although I did not test whether presentation of powerful/powerless genders (e.g., presenting gendered names paired with high/low status professions) would result in differently valenced ratings, Schubert (2005) found that his participants considered powerful/powerless groups as equally positive. Therefore, it is unlikely that thinking about *powerless* involved negative valence that affected decision times. Taken together, it seems that the powerless-bottom association may be less salient or accessible in cognition than the powerful-top association, but such differences are independent of valence.

Spatial simulations and implicit attitudes

In terms of the attitudes towards high-status males/females, participants had a tendency to associate males with high status and these associations were not related to spatial simulations of either males or females. Turning to the rationality/emotionality and gender IAT, consistent with my hypothesis participants held associations between males and rationality, but not between females and rationality. Such associations were not related to spatial simulations of either males or females. In contrast, the same was not true for top-rationality associations (inconsistent with Cian et al., 2015). Although

[32] Linguistically marked concepts refer only to themselves (i.e., sad), whilst unmarked concepts refer to the dimension as a whole (e.g., happy; Proctor & Cho, 2006)

on average, participants did not exhibit response latency facilitation to top-rationality items as opposed to top-emotionality items, I found that those who had a tendency to associate top locations with rationality were also faster to categorise males as powerful at the top in contrast to females as powerful in the same location. These findings suggest that having a more pronounced concrete representation of abstract categories (in this case rationality), involved facilitation to simulate males as powerful when they appeared at the top. However, as these results did not replicate in terms of social-status-males and rationality-males associations, it is difficult to draw strong conclusions about these findings.

The independence of spatial simulations and implicit attitudes might be attributed to the nature of the tasks I used to assess them. The spatial task required a certain degree of control in making a judgement – participants needed to explicitly think about powerful/powerless individuals when making a response in each trial. On the other hand, the IATs do not rely on any form of explicit thought, as in these tests, participants are required to quickly sort words into categories without engaging into a more deliberative thought. Hence, the differences in automaticity needed in performing each task could have obscured any potential associations between attitudes and spatial simulations.

Overall, it appears that spatial simulations of gender and power are independent of implicit attitudes. Yet, top-rationality association seemed to be related to response facilitation to males as powerful at the top as opposed to females, suggesting that more concrete associations between abstract thinking and verticality are associated with more pronounced spatial simulations of stereotype-consistent individuals.

Spatial simulations are moderated by context

I detected spatial simulations only in the context of thinking about males as powerful, but not females as powerless. Indeed, previous literature indicates that spatial simulations are not consistently detected across all contexts (e.g., Borghi, Glenberg, & Kaschak, 2004; Bub, Masson, & Cree, 2008; Lebois et al., 2015). My findings extend this perspective in a novel and important way by suggesting that an application of the power concept to social groups, and especially stereotypic associations about those groups (males-powerful), enhances the activation of spatial features of the grounded concept of power. This is based on the extent to which individuals hold an association between *male* and *power* which, as a general tendency, is implied by my finding that participants selected males as powerful more often than females. That is, the male-power link tends to be salient and guides participants' responses in the context of the choice task (see also Rudman & Kilianski, 2000). At the same time, the stereotypic associations do not always trigger spatial simulations, as participants did not simulate females as powerless at the bottom – even though their response pattern indicated a slight advantage (however, not significant one) at choosing females as powerless at the bottom as compared to males as powerless at bottom.

In relation to my previous arguments, I would like to note that my spatial reaction-time test was a forced-choice task in which participants were required to provide a response, even in situations when they might not have found either of the two individuals particularly powerful/powerless. Hence, random fluctuations in the salience of the stereotype might have taken place during the experiment across trials. The random fluctuations argument finds support in a recent review article on temporal instability of implicit biases. Payne, Vuletich, and Lundberg (2017) argue that situational factors rather than stable individual attitudes are more likely to determine the accessibility of stereotype-consistent content in people's minds that lead to stereotype-consistent responses (see Müller & Rothermund, 2004, for empirical evidence). Such reasoning seems logical in the context of research showing low test-retest correlations in individual levels of implicit prejudice in longitudinal studies (see Gawronski, Morrison, Phills, & Galdi, 2017). Also, social categories in cognition are typically abstract and may subsume variable sets of exemplars. Therefore, the context of thinking might affect accessibility of concepts and exemplars, and might be more important in determining which associations will be retrieved when making a particular judgment (Medin, 1989). Under such conditions, the context of my task possibly stimulated participants, in some trials, to activate the power concept in a counter-stereotypic situation (i.e., considering females as powerful), such that the spatial features of power, in such trials, may then have suggested a top-position for the female.

These suggestions are also consistent with research on multiple categorisations and goals activation during stereotypic judgements. The use of the gender stereotype might not be always consistent with participants' goals in a particular situation (e.g., comprehension of the situation has been achieved and participants are motivated to avoid prejudice; Kunda & Spencer, 2003). Also, when an alternative categorisation is available, participants would be more likely to inhibit the stereotype and therefore make a stereotype-inconsistent judgement (Hall and Crisp, 2005). The choice of the alternative categorisations could be further dependent on self-enhancement goals. Because the majority of my participants were female, it is possible that such counter-stereotypic responses were also associated with a motivation to demonstrate that females can also be powerful. Overall, I speculate that the mapping of abstract concepts onto the concrete dimension of space is particularly likely under situational factors that engender a high (stereotypical) consistency between the concepts involved.

However, because context was not directly manipulated in the presented studies (1 - 6), it is difficult to draw strong conclusions about contextual effects on spatial simulations. I address these issues in the next sections of the present chapter.

Also, it is worth noting that all of my studies are based mainly on samples of female participants. As noted earlier, the samples were representative of their participant panels. Previous research has indicated that females have equally negative implicit attitudes towards powerful women as males, but they are less explicitly prejudiced against females as powerful than males (Rudman & Kilianski, 2000; Rudman et al., 2012; Rudman, & Phelan, 2010). Although I strongly believe that spatial simulation in cognitive blending is likely to be a general phenomenon across gender (and it appears that it is independent of people's implicit attitudes, as found in my studies), it would be important to experimentally test whether males and females exhibit differential spatial simulations of males as powerful and females at the top directly.

Finally, I would like to note that the conclusions drawn in Chapter 2 are based on an integrative analysis across six experiments. As not all individual studies (see Appendix 6) achieved standard significance levels, it is important to acknowledge that while testing the same hypothesis across several studies, it is unlikely that each of the studies would support that prediction (see Lakens & Etz, 2017). Therefore, an integrative analysis provides more reliable results (Lakens & Etz, 2017). At the same time, the results obtained across six experiments were associated with small effect sizes. The indirect nature of the methodology used to identify spatial processes could be responsible for such small effect sizes, which were also found in other studies investigating spatial congruency effects (Schubert, 2005; von Hecker et al. 2013, 2016). Therefore, it is important to conduct more research on spatial congruency effects to establish that despite the small effect sizes, sensorimotor involvement in abstract thinking is a reliable process.

An eye-opening measure of stereotyping - Chapter 3

In Chapter 3, I sought to replicate my previous findings by incorporating a pupillometry measure. My main motivation to use such method was to replicate previous findings with a physiological technique. This can be achieved with pupil size measure that can be used as a proxy of experiencing surprise and also cognitive load. First, I wanted to examine whether the mere presentation of gender in *incongruent* spatial positions (males-bottom; females-top) would be associated with initially increased arousal due to experience of surprise. Second, I wanted to test whether such cognitive conflict and a subsequent attempt to control the automatically activated stereotype of males as powerful would also involve increased cognitive effort after people had made the response. To address the above issues, I first investigated whether differences in pupil size can be used to detect the basic associations between groups with clear power implications and spatial positions by replicating Schubert's (2005) work. Second, I investigated whether this method could be also applied to an investigation of the mental representation of gender stereotypes.

Schubert block

Overall, consistent with my hypothesis, I found that across the whole trial period, participants exhibited increased pupil size on trials where they were presented with powerful groups at the bottom as opposed to the top. Such findings suggest that participants experienced a higher cognitive conflict when they needed to process powerful groups at the bottom. In the post-response period, I found that this difference was significant immediately after participants made their responses (i.e., in the period from 1379ms to 2000ms). The differences most likely reflect increased cognitive effort (see Kahneman & Beatty, 1966).

The response-locked pupil dilation is consistently observed in studies employing cognitive-conflict paradigms such as Stroop task or flanker task (Geva et al., 2013; Laeng et al., 2011) demonstrating the pupil dilation peaks on incongruent trials. It is proposed that such peak dilations reflect inhibition of distractors or attempts to control the spontaneous response tendencies, background noise, and response evaluation (Laeng, et al., 2011; van Steenbergen & Band, 2013). In general, such dilations can be attributed to increased cognitive effort on incongruent versus congruent trials (van der Wel & Steenbergen, 2018). Therefore, as demonstrated in my study, the power-bottom incongruence also involves increased pupil dilation that occurs after participants decided that the powerful group was presented at the bottom.

Further, the literature on pupillary correlates of cognitive effort (Geva et al., 2013, Laeng et al., 2011; van der Wel & Steenbergen, 2018) suggests that the differences in pupil sizes across congruent and incongruent trials are mostly observed *after* participants make responses. This is because when the decision is executed, the LC-NE system is activated in the phasic mode, which is one of the firing modes of LC neurons associated with behavioural exploitation (Aston-Jones & Cohen, 2005). The phasic LC firing is involved in task engagement (as opposed to the tonic mode associated with exploration or seeking alternatives). Further, task-engagement associated with LC-NE phasic activity is proposed to facilitate decision-making and ensure that the chances of completing the task successfully are optimised. However, my results also indicate that there was a significant difference in participants' pupil size between trials where powerful groups were presented at the bottom versus top in the *pre-response* period. Similarly, I observed increased pupil size on trials where participants were presented with powerful groups at the bottom versus top at 750ms time stamp. Therefore, it is likely that these differences can be attributed to initial perception of inconsistent information and cognitive conflict. Such an explanation of my findings is in line with research indicating that early surprise evoked by error prediction, unconscious perception of anomalous playing cards or modified faces (i.e., Thatcherised faces), correlates with increased pupil dilation (Preuschoff et al., 2011; Proulx et al., 2017; Sleegers et al., 2015). In addition, Smallwood et al. (2011) reported that pupil dilation is higher in the pre-response period on trials where participants give an incorrect response to a working-memory task. In their task, participants were presented with numbers and were supposed to indicate whether the numbers were even or odd. In contrast, in the Schubert block, I observed such differences when participants answered *correctly*, that is, when they indicated powerful groups at the bottom. Such findings suggest that the incongruent spatial location was involved in the increased pupil dilation in my study. Yet, when Laeng et al. (2011) investigated pupillary responses in Stroop task trials, they noted that pupil size was increased at pre-response periods, however, this dilation occurred to the same extent across congruent and incongruent trials. The only differences in pupil dilation were observed later in the post-response period. In principle, both Stroop and Schubert blocks should reflect initial cognitive conflict reflected in pupil dilation, but it is essential to note that the tasks are qualitatively different from each other and might employ different cognitive processes. Schubert block responses are based on *social* knowledge rather than inherently true or false answers like naming colours.

To conclude, the most likely explanation for increased pupil size in the pre- and post-response period in my replication of Schubert's task is that cognitive conflict and load, respectively, were higher when participants needed to make a judgement about powerful groups presented in the incongruent spatial positions. It is also plausible that in general, the initial increase in pupil size on incongruent trials (powerful-bottom) was associated with cognitive conflict that in turn involved higher cognitive effort in decision-making.

Gender block

Consistent with my hypotheses, I found that participants' pupil size was associated with an interaction between gender choice and trial type. Within the whole response period, I observed increased pupil size on trials when males were selected as powerful at the bottom as opposed to the top - supporting my findings from my previous studies and the Schubert block. The same results were obtained when the post-response period was analysed separately.

Second, as hypothesised, participants did not experience higher cognitive effort when selecting females as powerful when their names appeared at the top or bottom. Overall, such findings corroborate my reaction-time data obtained in Studies 1 – 6. These studies demonstrated spatial simulations of males as powerful only. In particular, in the present study, the pupil size data indicate higher cognitive effort associated with processing of and choosing males names at the bottom while selecting the powerful person between two gendered names available. These findings therefore suggest that attributing the concept of power to stereotype-consistent stimuli is associated with spatial simulations. That is, the mental representation associated with stereotypic thinking is more concrete.

Third, when analysing the entire-trial period, I found that participants exhibited marginally higher pupil dilation when they selected female names when they appeared at the top as opposed to male names at the top. Again, this difference supports my hypotheses and previous findings from Studies 1 – 6, suggesting higher cognitive effort when processing stereotype-inconsistent stimuli at the top (i.e., powerful-females). These findings are interesting, as it appears that choices of stereotype-consistent (powerful-males) and –inconsistent (powerful-females) targets per se do not involve higher pupil dilation. Instead, mental effort observed in my study was specifically associated with spatial location interacting with stereotypic thinking.

My analyses of the pre-response period suggest that participants experienced surprise when they were presented with trials on which female names appeared at the top and male names at the bottom. This is because they demonstrated increased pupil size when presented with trials where female names appeared at the top as opposed to male names in the same location and such pupil size dilation was *independent* of participants' response later on within the trial. The results are in line with the findings obtained in the Schubert block showing that participants also experienced increased pupil size in the pre-response period on trials presenting powerful groups at the bottom as opposed to the top. Interestingly, such differences were noticed at 750ms in both blocks. Because the differences were independent of participants' later decisions about males as powerful or females in the gender block, it appears that higher arousal associated with observing female names presented at the top (in the context of looking for a powerful person) was associated with increased cognitive conflict. Such cognitive conflict seemed to be then responsible for the increased pupil size in the early period of the trial (Lin et al., 2018; Preuschoff et al., 2011; Proulx et al., 2017; Sleegers et al., 2015).

Furthermore, it is important to note that the findings of the marginal difference observed in pupil size between choices of females as powerful at the top as opposed to males as powerful at the top were not replicated when post-response period was analysed separately. This is surprising because processing of inconsistent information that requires inhibition of the interfering distractor should be associated with LC-NE phasic activity and therefore especially increased pupil size (Aston-Jones & Cohen, 2005; Geva et al., 2013; Laeng et al., 2011). It is possible that the general increase in arousal in the entire trial within my study was primarily driven by the initial cognitive conflict of perceiving females at the top while tasked with selecting the powerful person. Such results could be also attributed to the fact that participants could not compare both males and females at the top within the same trial. Therefore, processing of females at the top would be mainly associated with distinguishing them from males at the bottom.

In fact, my additional analyses of differences between choices of males and females in different trial types indicated that participants exhibited marginally higher cognitive load on trials when they picked female names as powerful at the top as opposed to male names at the bottom. Interestingly, the increased pupil size to the presentation of female names at the top at 500ms, but not 750ms, was marginally associated with later categorisations of those females as powerful. Further, it can be speculated that the increased pupil size at this early stage, which was most likely associated with aversive unconscious arousal to an inconsistent stimulus (females at the top), stimulated participants to engage in compensatory behaviours and hence counter-stereotypic choice. Such findings are in line with suggestions by Proulx, Inzlicht, and Harmon-Jones (2012), who argue that aversive arousal associated with unexpected stimuli is compensated by either accommodation or assimilation of counter-attitudinal or stereotypic information. However, these results are marginal and they do not imply any causal link between increased pupil size and later categorisations of female names as powerful at the top. Therefore, these suggestions should be treated with caution.

Dependency of spatial simulations on task features - Chapter 4

In Chapter 4, I considered the precise mechanisms that were likely driving spatial simulations detected in Chapters 2 and 3. First, I designed Studies 8 and 9 in order to test whether verticality cues per se, that is, the spatial location (top or bottom) of stimuli, would lead to a response facilitation when people make a judgement about power differences, regardless of the semantics of the stimuli presented. Specifically, I investigated whether verticality-power associations are always present within spatial simulations. That is, I examined whether any concept could be associated or blended with power and whether such blending would involve spatial simulations (see Schubert, 2005). To achieve this, I asked participants to quickly detect either the powerful or powerless person on the screen between two female or two male names. In this way, males and females were not directly compared in each judgement.

Second, while in Studies 8 and 9, I explored the independent role of verticality cues when participants did not directly compare power differences between males and females, in Study 10, I aimed at investigating if gender per se can be spatially simulated when the power context is not salient. To test that, I asked participants to quickly identify whether a vertically presented name (top versus bottom) was either male or female.

The findings from Study 8 indicate that spatial locations, power, and gender stereotypes are not associated when participants compare either pairs of female or male names, even when they are matched with high social-status professions. This indicates that verticality cues per se do not play a role in facilitating participants' judgements. Therefore, verticality cues are not sufficient to evoke spatial simulations of power in the absence of strong power differences between compared individuals. Further, I found a facilitation to the powerful condition. Participants were faster at categorising male and female names as powerful versus powerless. Such findings are in line with Schubert's (2005) results that also indicated a processing advantage to the powerful condition. Therefore, these findings might suggest that in the absence of salient stereotypic associations, the linguistic markedness of the word *powerful* results in a significant processing advantage (Hamilton & Deese, 1971).

Finally, I found that there was a higher marginal tendency to select top positions than bottom ones when presented with both male and female name pairs. However, such results were true for participants selecting powerful and powerless names. It is likely that the reading habit (from the top to the bottom) was responsible for such effects.

Further, in contrast to Study 8, where I presented just gendered names, in Study 9, I found a facilitation of response times to female targets versus male targets. This was unexpected, as across all previous studies, I did not detect consistent effects of gender per se on response times. However, it is likely that such findings are associated with in-group effects – most participants were female, so it is possible that they responded faster to the names associated with their own gender. In addition, I also found that participants were significantly faster to select female names as powerless as opposed to male names as powerless. It is difficult to draw strong conclusions on the basis of these findings, as no complementary effects were detected in the powerful condition. In terms of proportions of choices, participants were more likely to pick names as powerful at the top and as powerless at the bottom, which is in line with the majority of my previous studies. Yet, this effect was moderated by gender block. Powerful-top and powerless-bottom associations were only present in the case of male targets and not the female ones. Therefore, it appears that verticality cues guided participants' *types* of responses.

To summarise, across Studies 8 and 9, I found that when a power context is not made salient, spatial cues do not have an effect on response latencies when participants make judgements about power. Similarly, in Study 8, I found that both male and female targets were categorised faster in the powerful condition compared to the powerless one in both spatial locations. Therefore, there was a general processing advantage when participants selected *powerful* individuals as opposed to *powerless* one, and so this seems to be independent of spatial locations. Furthermore, such findings were not replicated in Study 9 and this is probably because participants were presented just with gendered names and without high-status professions. This is because high-status professions provided more a power-relevant context in contrast to a sole presentation of names. Overall, the results point to the idea that attributing social power in a power-relevant context is associated with spatial correlates. Yet, when the context is ambiguous (social status or gender stereotype are not made salient), people do not experience a processing advantage while selecting individuals at the top or at the bottom.

Nevertheless, across two studies, I found consistent evidence for spatial correlates in the gender-power mental representation in terms of the frequency with which participants selected males/females at the top/bottom as powerful or powerless. Although there was an overall tendency to pick the top names as powerful and the bottom name as powerless, when gendered blocks were analysed separately, participants were more likely to pick male targets as powerful at the top versus bottom (and vice versa for the powerless), however, this was not the case for female targets. Such results are partially consistent with my previous studies (1 - 7), suggesting that spatial features of power are activated only in stereotype-consistent categorisations, specifically, males as powerful. However, I also observed a higher tendency to pick males as powerless at the bottom versus the top, suggesting that concrete thinking about both powerful and powerless individuals is associated more with representing male, but not female targets.

Altogether, it seems that comparing power differences between the same gender names (females-females/males-males) does not involve spatial simulations, as reflected in the lack of response facilitation. However, the likelihood with which people categorise male, but not female, names as powerful and powerless involves verticality cues regardless of stereotype-fit. This occurs even when participants do not directly compare power differences between male and female individuals.

Turning to Study 10, I found that semantics of gender are not directly associated with the vertical spatial dimension. The results suggest that mapping of gender onto the vertical dimension involves complex associations between semantics of gender, spatial locations, and finally stereotypic associations. Therefore, to be simulated in space, gender requires a meaningful power context that could facilitate its spatial mapping. That is, only when power is blended with gender in the power-relevant context, spatial features of power are activated. It is likely that such process is then associated with simulating gender on the vertical dimension.

The results from Study 10 also showed that the interaction between vertical position and response mode was not significant contradicting the polarity correspondence theory (see Proctor and Cho, 2015, for an overview). Weeks and Proctor (1990) demonstrated that people respond faster to a stimulus presented at the top with a right key, such as L, and a stimulus presented at the bottom with a left key, such as A. This is because up and right spatial locations linguistically represent an unmarked end of the dimension, so they are +coded (i.e., they have a positive polarity) in contrast to bottom and left locations that are -coded (i.e., they represent a negative polarity). Finally, if polarities of spatial locations of stimuli are matched with the polarities of the response mode, then a facilitation in reaction-times to such stimuli would occur. This was not supported by my data.

Spatial simulations and stereotype accessibility - Chapter 5

In Chapter 5, across two studies, I explored the role of stereotype activation in the mental representation of power and gender on the vertical dimension. Consistent with my hypotheses, in Study 11, I found that gender has power implications and these implications are more pronounced when people are given more processing time. When determining the powerful person between opposite gender names, participants picked males as powerful more often than females. This effect was especially prominent in the long condition (2000ms to respond) compared to the short one (600ms). I also demonstrate that both male and female names were selected as powerful more often at the top than at the bottom and this effect was equal for both genders, even though males are more associated with power than females. Such findings show that stereotypes do not moderate top-power associations.

Integrating the presented findings with the results from my previous studies, it appears that stereotype-consistent thinking is only essential in *spatial simulations*. That is, stereotype activation seems to be associated with the duration with each people process or think about certain concepts. At the same time, stereotype activation becomes irrelevant when it comes to the *outcomes* of participants' judgements. That is, in the current study, in which participants did not give speeded judgements, verticality cues guided their responses regardless of whether the chosen individuals conformed to the gender stereotype or not. It is likely that spatial simulations interact with stereotype activation only when people need to decide as fast as possible which person was powerful, but not in other circumstances.

Finally, in Study 12, I did not detect any of the hypothesised effects. The study was underpowered (only .10 of power was achieved) due to a small sample size – also some participants needed to be removed from the analysis, as they knew the purpose of the study or were aware of the powerful-females/powerless-males pairings. Therefore, it is difficult to draw strong conclusions on the basis of that study. So far, the results indicate that priming powerful-females and powerless-males does not have implications for how often people select male and female names as powerful. In this study, I also did not replicate spatial simulation effects (a facilitation to selecting male, as opposed to female names as powerful when they appeared at the top; and selecting males as powerful faster when they were presented at the top versus bottom).

According to grounded cognition theory, abstract concepts like power are grounded in sensorimotor experiences, and their cognitive representations involve properties derived from these experiences (Barsalou, 1999). The literature clearly indicates that power is associated with the vertical spatial dimension (Schubert, 2005). In Study 11, I extended this line of inquiry by demonstrating that the spatial features of power are activated in the mental representation of gender, such that males and females considered as powerful are associated with the top vertical position. These findings are in line with Schubert's (2005) results, further indicating that in the meaningful power-context, spatial features of power (top and bottom vertical positions) play a role in the mental representation of the concept that is attributed to power. However, it is important to note that these findings are based on the frequency with which participants indicated each gender in each vertical position. That is, participants' *choices*, or in other words, the outcome of the decision-making was associated with upper positions. I consider this process (associating verticality with power as demonstrated by the frequency of choices) as being not directly linked with spatial simulation. This is because spatial simulation is an active process that could be captured by measuring the *speed* of processing, but not the outcome of the thought process.

Further, I found that even though participants selected males as powerful more often than females (especially in the long condition), stereotype-fit (considering males as powerful) per se did not moderate the power-space associations – any name (male or female) picked as powerful involved a similar (and greater) tendency to choose them at the top versus bottom. Hence, making a judgment about gender (which carries latent power implications) in an explicit power-related context activates spatial correlates of power leading to representations of that concept on the vertical dimension.

The impact of spatial vertical presentation on gender perceptions - Chapter 6

In the last empirical chapter of my thesis, I aimed at investigating whether the previously detected spatial correlates in the mental representation of gender stereotypes might bias people's perception of gender in stereotype-consistent ways. Specifically, I wanted to examine whether the concrete spatial representation associated with mental representation of gender stereotypes might involve reinforcement of stereotype-consistent judgements. In two studies, I demonstrated that the vertical presentation of male and female targets can bias people's subsequent perceptions of gender. Study 13 shows that vertical presentation in the context of gender stereotypes produces biased perception of size. Study 14 further demonstrates that such biasing effect is not only confined to concrete judgements, but this effect extends to abstract judgements about warmth and competence.

Overall, in Study 13, I found that people have a biased perception of gender in terms of physical size. Participants tended to overestimate the size of male names as opposed to female names regardless of the condition. I also found that when participants were primed with stereotypically-consistent persons (i.e., males) at the top (when a subtle power context was activated), they overestimated (underestimated less) the sizes of subsequently presented both male and female names. This effect was the most pronounced in the case of males-top as opposed to females-top priming or control conditions. In general, my findings indicate that the mental representation of power involves spatial features in the context of gender representations and such spatial features bias perceptions of physical size. This effect is the most pronounced when priming people with the concept of males positioned at the top. Therefore, it seems that the presentation of males on top has a *spillover* effect to the gender other than the most relevant one (i.e., males). Future research should investigate whether such effects would be also extended to other social/non-social categories. It is also worth noting that the stimuli presented (i.e., males-top and females-top trials) in my task might not act as prime per se. It is more likely that congruence between a mental simulation associated with the stimulus and the spatial arrangement of a stimulus presentation is responsible for the revealed effects (i.e., more overestimation in the males-top versus control condition). Specifically, it is possible that the interaction of a stimulus (e.g., male names presented at the top) with its pre-existing mental representation (males are stereotypically more powerful than females and power is associated with upper spatial location) constitutes the prime. However, this speculation needs further empirical support, as my data cannot indicate precisely what kind of mental representation was associated with the detected effects.

According to grounded cognition, abstract concepts like power are grounded in sensorimotor experiences, and their cognitive representations involve properties derived from these experiences (Barsalou, 1999). The literature clearly indicates that power is associated with the vertical spatial dimension (Schubert, 2005). As males were seen as more powerful and metaphorically power correlates with physical size (Lakoff & Johnson, 1999; Schubert, 2005), in the present study I found that participants were more likely to overestimate the size of male names relative to female names, an effect that was consistent across conditions. Interestingly, the presentation of male names at the top, without any explicit statement suggesting that males were powerful, primed perceptions of greater sizes of any subsequently presented stimuli. This effect was present, but smaller, when female names were presented at the top. The results suggest that the upper vertical presentation of a social group which has implicit power implications biases perceptions of physical sizes stimulating people to see subsequently presented category as bigger than it is in reality.

While Study 13 presents evidence that location can bias size perception, in Study 14, I aimed at testing whether spatial simulations (quicker response to stereotype-consistent person in the congruent spatial location) would also bias stereotype-consistent perceptions of gender in terms two stereotype dimensions associated with SCM, i.e., warmth and competence (Fiske et al., 2002). First, by using pictures instead of gendered names, I detected spatial simulations of males as powerful and females at the top versus bottom. This is in line with the findings from Study 11, which suggested that mental representations of power involve spatial features regardless of whether the power concept is attributed to male or female targets. The results from Study 11, however, were only based on the *outcomes* of participants' thought processes, that is, the frequency with which they chose each gender in each vertical position. Recall, such choices are different to reaction times, e.g., the shorter the time to categorise males as powerful, the higher the ease to process such stimuli. This distinction is important, as by *simulation* I refer to the processing time (hence the effort) with which participants make their decisions, whilst the *outcome* or *choice* indicates the type of judgement people made. Therefore, the current findings indicating that both males and females are *simulated* at the top show that participants found it easier to process both genders as powerful at the top as opposed to the bottom. This is not in line with my integrative analysis (Studies 1 - 6) and the pupillometry evidence from Study 7 that indicated processing advantage for males as powerful at the top, but not females. However, it is important to note that when participants were presented with gendered names and professions in Study 3 only, they also showed evidence of spatial simulation of females as powerful at the top versus bottom. The most plausible explanation for such findings is that in the present experiment, I used pictures as stimuli in contrast to gendered names. It is possible that the vividness of participants' associations between females and power was increased, which was related with increased salience of top vertical positions. Increased salience of the vertical positions could be then associated with more readily accessible links between power and verticality resulting in representation of females at the top. In Study 3, such an effect could be attributed to the high-status professions that were presented with gendered names.

In terms of the judgement task, participants tended to dislike males more when they simulated them spatially at the top. This might demonstrate a rejection of the stereotype, especially given that my sample was mainly composed of female participants. Further, according to SCM (Fiske et al., 2002) people should perceive females as high on warmth and low on competence, whilst the opposite should be true for males. I predicted that such perceptions should be especially reinforced by spatial location of simulated individuals. These predictions were not supported by my data.

Although the simulation scores were not associated with warmth and competence judgements, I found that the likelihood of categorising genders in different spatial locations was significantly correlated with perceived levels of warmth and competence. Consistent with SCM, spatial correlates present in the mental representation of females as powerful were associated with lower femininity ratings of female targets. Interestingly, this correlation was only true for choices of females at the top and not at the bottom. Such findings are in line with the literature on ambivalent sexism and gender roles suggesting that femininity is stereotypically associated with warmth and lack of competence (Glick & Fiske, 1996; Rudman & Kilianksi, 2000; Rudman & Phelan, 2008). Indeed, SCM points out that career women are perceived as competent, but not warm, as they do not conform to the traditional views of women of being warm and incompetent. As the found association between femininity and females as powerful was specific only to the top positions, it seems that spatial location plays a role in reinforcing gender stereotypes. Consistent with these findings, when males were picked as powerless, but only at the bottom, they were perceived as less masculine and competent. Again, these results are in line with gender stereotypes, as stereotypically, masculinity is associated with competence and dominance. The bottom spatial location seems to reinforce such beliefs. Finally, spatial location was not relevant to participants' judgements when they categorised females as powerless, as they considered them as higher on warmth regardless of the indicated position. Although these results are in line with SCM and suggest that spatial locations indeed reinforce stereotype consistent views, some of the predicted correlations were not detected. For example, in principle, the likelihood of choosing males as powerful at the top should be associated with their higher competence and masculinity ratings. Further studies should investigate under what circumstances such associations might emerge.

Alternative theoretical interpretations

I wish to argue that conceptual congruency effects (or spatial simulations in conceptual blending) were responsible for participants' faster responses to males as powerful rather than females as powerful at the top in my initial studies (Barsalou, 1999). However, it is possible that other processes have contributed to my observed pattern of results. Polarity correspondence theory (Proctor & Cho, 2006) postulates that some spatial facilitation effects occur because of the correspondence in polarities between concepts and people's responses to those concepts. That is, +polar (i.e., unmarked concepts) and -polar concepts (i.e., marked concepts) can be matched with +polar/-polar responses. If the matching is congruent (e.g., +polar concepts are matched with +polar responses), then processing will be facilitated. Therefore, this theory predicts that structural features of concepts in terms of polarity (i.e., linguistic markedness: powerful and top responses would be +polar; whilst powerless and bottom -polar) are enough to produce facilitations of responses to powerful individuals at the top and powerless at the bottom. This prediction is partially supported in my studies – participants were quicker to detect males and females as powerful at the top of the screen. On the other hand, I did not find significant processing advantages in the case of participants' responses to choices of males and females as powerless at the bottom, which polarity correspondence would predict. In addition, it is not clear how gender should be coded in terms of polarity. Therefore, spatial congruency effects of power interacting with linguistic markedness in the case of thinking about gender is a more plausible explanation for my findings. Overall, the assumptions of polarity correspondence cannot fully account for my results (i.e., the role of stereotypic thinking). Further research is necessary to address these theoretical issues.

Limitations

Spatial task

I would like to point out that my findings are based on a quasi-experimental design in which gender choice was not manipulated per se, but was, as a variable, self-generated by participants. I firmly believe that this approach is important as it allows participants to generate their own responses about stereotypicality, rather than being constrained to make responses that would be correct or incorrect (as in Schubert, 2005). At any rate, my results suggest an *association* between stereotypic thinking and spatial simulation rather than a causal link. As I argued earlier, spatial simulation is a process that likely underpins or supports thinking. It is possible that concrete spatial reasoning about abstract concepts serves to facilitate judgments either by making people more subjectively confident about their decisions or by increasing the efficiency of their decision-making. Yet, further empirical evidence is needed to elucidate such questions. In future research, a direct manipulation of stereotypic consistency could bring more insight into the potential causal link between thinking about gender and spatial processing.

Another issue associated with the quasi-experimental design was that the number of trials on which participants indicated each of the choice categories (males as powerful/powerless and females as powerful/powerless at the top or bottom) could not be standardised. That is, the trial number for each choice category was unequal. Although I found that the proportion of participants' choices was unrelated to their reaction times, future studies should address this limitation by employing a more standardised paradigm.

Participants

I note that most of my participants were female. This might have implications for interpreting the results. On the one hand, females might perceive gender-power links differently than males, as traits of out-group members are usually seen as more homogenous (see Mason, 2006). Also, my participants were more likely to respond in stereotype-consistent manner, especially when they had more time to think about the power concept in the context of gender. This indicates that the relatively diminished cognitive load does not stimulate counter-stereotypic judgments, as females might be less motivated to appear less "prejudiced" (Fiske et al., 1999; Macrae et al., 1997). On the other hand, a study based on a large dataset indicated that participants' gender is only a minimal moderator of sexist beliefs, as sexism is best predicted by political attitudes (Roets, van Hiel, & Dhont, 2012). Similarly, Rudman and Kilianski (2000) found that both males and females have equally negative implicit associations about females who have a high social status. Overall, the evidence seems to indicate that people's gender might not in fact moderate stereotypic associations or override learned social norms. Yet, future research should investigate spatial simulations in the context of gender stereotypes using also mainly male participants.

Pupillometry

It is also important to note that conclusions of Chapter 3 are based on an innovative measure of cognitive processing. Pupillometry has not been widely used within social cognition. Also, pupillary activity is an *indirect* measure of cognitive processes and LC-NE activity. Therefore, my findings should be treated with caution, as we need more empirical evidence to support presented results. Finally, raw pupil signal requires pre-processing. The level of pre-processing that is needed, however, varies depending on the eye-tracker used and quality of the signal obtained. Because of such variability, the techniques used to pre-process pupil signal vary across the literature (Cavanagh et al., 2014; Lin et al., 2018; Mathot et al., 2018). Hence, the lack of consistency in cleaning pupil data may compromise our ability to compare the studies reported within the literature. Nevertheless, as demonstrated by my data, the pre-processing techniques might affect the significance of the tested differences. On the other hand, it is unlikely that filtering/blink corrections/interpolations would qualitatively change the pattern of pupillary reactivity. Also, in the case of Study 7, my pupil size data mapped onto my previous findings, suggesting that pupillometry can be indeed used to investigate cognitive processes in the context of social phenomena.

Implications

Overall, my findings indicate that the perceptual symbol of space associated with sensorimotor experiences is present when people make stereotype-consistent judgements about gender and power. That is, stereotype-consistent decision making involves concrete mental representations, whereby males, but not females, are simulated in the upper vertical positions when considered as powerful. The literature on stereotyping indicates that the accessibility of learned stereotypic associations in a given situation is an important factor in determining whether the stereotype will be activated and applied (Gilbert & Hixon, 1991; Kawakami, Dovidio, Moll, Hermsen, & Russin, 2000; Lepore & Brown, 1997; Sinclair & Kunda, 1999). Based on my findings, I speculate that the spatial features associated with the mental representation of gender stereotypes might promote such accessibility by increasing the salience of stereotypic links (males-powerful). Such links might in turn be associated with an increased tendency to apply gender stereotypes in situations where power is relevant. For example, when evaluating employees' work performance, considering them for promotion, or choosing appropriate job candidates, people might be prone to implicitly apply the gender stereotype due to its concrete and therefore more vivid spatial representation. This might be especially true when the employees' or candidates' gender is salient, that is, when females are compared to males (Ford & Stangor, 1992).

Final conclusions

In Chapter 2, I demonstrated that cognitive blending between social categories and power is assisted by spatial simulations and this process is dependent on stereotypic beliefs. When people reason about power in the context of gender categories, they exhibit pronounced spatial simulation effects when they think about those categories as powerful, and this occurs in the case of stereotypic associations (i.e., males as powerful). I argue that blending the concept of power with social concepts that have no immediate power implications (i.e., gender should be in principle power-neutral) is sufficient to activate the spatial features of grounded concepts, which are then used in spatial simulations. Therefore, cognitive blending in human cognition depends on both the compatibility between mappings of space and abstract social concepts as well as socially constructed stereotypical beliefs. These are the main aspects of concepts that compete for activation within a mental model to accomplish conceptual integration and meaning production.

The above findings were replicated with pupillometry. I found that across the two blocks of trials (the Schubert and gender block), it seems that people experience higher cognitive effort when they indicate powerful groups and males as powerful at the bottom as opposed to the top. Further, males as powerful seem to be processed with more ease when they are presented at the top and bottom than females as powerful at the top. Such results are based on pupil size data indicating higher pupil dilation on the incongruent Schubert trials (powerful-bottom) as well as when participants indicate a counter-stereotypic individual as powerful and when that choice interferes with spatial positions (powerful-females-top versus powerful-males-top and -bottom). The cognitive effort account can in principle explain the findings of increased pupil size after participants made their responses in my task. This is because it accurately accounts for my finding in the light of the recent review on pupillometry presented by van der Wel & Steenbergen (2018) pointing out that pupil size is directly associated with cognitive effort. Furthermore, increased pupil size in the pre-response period could be associated with cognitive conflict involved in preceptions of counter-stereotypic or incongruent information (Preuschoff et al., 2011; Proulx et al., 2017; Sleegers et al., 2015; Smallwood et al., 2011).

Having established the robustness of spatial simulations in stereotypic associations, in the next chapter I explored the dependency of spatial simulations on task features and stereotype activation. In Chapter 4, I found that when power is attributed to male and female names in the absence of an explicit power context (i.e., power differences are made between same-gender names), spatial simulations do not occur. Further, I also demonstrated that gender per se is not associated with spatial simulations. To conclude, the mechanisms driving spatial simulations observed in my previous studies (1 - 7) are likely to be based on the spatial correlates involved in the mental representation of power - spatial features can be observed in abstract thought only when the concept of power is activated in relevant conditions. This occurs when thinking about power is meaningful due to stereotypic associations (Studies 1 - 7). This is because when the power context is not present, spatial processing becomes irrelevant and hence it is not found (Studies 8 and 9). This is also in line with conceptual blending, as spatial simulations are likely to occur when two concepts are combined in semantically consistent manner (Fauconnier & Turner, 1998).

In Chapter 5, I demonstrated that cognitive components of social power involve sensorimotor features associated with verticality in the type of decisions that people make. Power seems to be associated with verticality to the same extent whenever it is attributed to male or female targets. At the same time, in my previous studies focusing on the *duration* (not the *outcome*) of the thought processes associated with linking gender-power and spatial location, I found that stereotype-consistent thinking is important in spatial simulations (Studies 1-7), as males as powerful, but not females were simulated at the top. Therefore, it is likely that spatial simulations facilitate judgements about social power of males and females, only when people make stereotype-consistent choices. However, I further found that the mental representation of power always involves spatial components, regardless of the social category it is attributed to (powerful-males/females as powerful) and such spatial components are associated with the *outcome* of judgements, but not *processes* leading to such judgements.

Finally, in the last empirical chapter of my thesis, I demonstrated that spatial correlates observed in mental representation of power and gender not only are present in cognition, but they also have a potentially reinforcing effect on stereotype-consistent judgements. In Study 13, presenting a stereotypically fitting person at the top of a display was associated with greater perceived size of subsequently presented male and female names. In Study 14, the likelihood of choosing females as powerful when they appeared at the top involved lower ratings of their femininity, whilst the likelihood of indicating males as powerless at the bottom was associated with lower ratings of their masculinity and competence. Overall, it appears that space is involved in abstract thought and possibly it reinforces stereotypic associations by providing a more concrete mental representation of stereotyped individuals.

Integrating the results across 14 studies, I showed that the mental representation of stereotypes about gender and power involves sensorimotor correlates. Such sensorimotor features are active when people reason about stereotypically-consistent powerful individuals, such as males, but not females. This is only true in power-relevant as well as semantically and metaphorically meaningful contexts (thinking about power differences between two individuals of opposite genders). These findings can be explained by the theory of conceptual blending, which suggests that mental representations of concepts that can be blended together, that is, combined in a semantically meaningful way, involve sensorimotor simulations. In turn, such an account is consistent with grounded cognition, proposing that abstract concepts (e.g., power) are conflated with more concrete concepts (i.e., space). Interestingly, the tendency to base power categorisation of gender on verticality cues is independent of the processing time with which such categorisations are made. Further, the type of power-gender judgment that is made seems to be associated with spatial features and such an association is independent of stereotype activation. Finally, spatial locations interacting with power-gender categorisations seem to promote stereotypically-consistent thinking suggesting that such spatial locations might in fact reinforce stereotypic mental representations.

To conclude, sensorimotor spatial simulations are involved in the mental representation of gender via perceptions of power. Individuals who are judged as powerful on the basis of stereotypes are represented in cognition in a more concrete way. Such spatial correlates seem to also bias perceptions of gender potentially reinforcing the concrete representation of stereotypes. It is possible that such concrete mental representations are more difficult to reduce. This should be investigated in future research.

References

Ahern, S., & Beatty, J. (1979). Pupillary responses during information processing vary with Scholastic Aptitude Test scores. *Science, 205*, 1289–92.

Allport, G. (1954). *The nature of prejudice.* London: Addison-Wesley.

Ambady, N., Paik, S. K., Steele, J., Owen-Smith, A., & Mitchell, J. P. (2004). Deflecting negative self-relevant stereotype activation: The effects of individuation. *Journal of Experimental Social Psychology, 40,* 401-408

Aston-Jones, G. & Cohen, D. J. (2005). An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance, *Annual Review of Neuroscience, 28*, 403-450.

Bargh, J.A. (2006). What have I been priming all these years? On the development, mechanisms, and ecology of nonconscious social behavior. *European Journal of Social Psychology*, *36*, 147–68.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22,* 577–609.

Barsalou, L.W. (2003). Situated simulation in the human conceptual system. *Language & Cognitive Processes, 18,* 513–62.

Barsalou, L.W. (2008). Grounded cognition. *Annual Review of Psychology, 59*, 617-645.

Bennett, M. & Sani, F. (2006). Contextual variation in stereotype content: An investigation of children's central tendency and group variability judgements. *Social Development, 15,* 692-708.

Berman, J. (2015). Millennial women are more educated than men, but are still paid less. *Market Watch*, Retrieved August 15, 2018, from: https://www.marketwatch.com/story/millennial-women-are-more-educated-than-men-but-are-still-paid-less-2015-04-08

Borghi, A. M., Glenberg, A. M., & Kaschak, M. P. (2004). Putting words in perspective. *Memory & Cognition*, *32*, 863–873.

Boroditsky, L., & Ramscar, M. (2002). The roles of body and mind in abstract thought. *Psychological Science, 13,* 185-189.

Brodeur, M. B., Dionne-Dostie, E., Montreuil, T., & Lepage, *M*. (2010). The bank of standardized stimuli (BOSS), a new set of 480 normative photos of objects to be used as visual stimuli in cognitive research. *PloS ONE*, *5*, e10773.

Bub, D.N., Masson, M.E.J., & Cree, G.S. (2008). Evocation of functional and volumetric gestural knowledge by objects and words. *Cognition, 106,* 27-58.

Carli, L. L., & Eagly, A. H. (2001). Gender, hierarchy, and leadership: An introduction. *Journal of Social Issues, 57,* 629-636.

Casasanto, D. (2009). Embodiment of abstract concepts: Good and bad in right- and left-handers. *Journal of Experimental Psychology: General*, *138*, 352-367.

Cavanagh, J.F., Wiecki, T.V., Kochar, A., & Frank, M.J. (2014). Eye tracking and pupillometry are indicators of dissociable latent decision processes. *Journal of Experimental Psychology: General*, *143*, 1476–1488.

Chiao, J. Y., Harada, T., Oby, E. R., Li, Z., Parrish, T., & Bridge, D. J. (2009). Neural representations of social status hierarchy in human inferior parietal cortex. *Neuropsychologia*, *47*, 354-363.

Cian, L., Krishna, A., & Schwarz, N. (2015). Positioning rationality and emotion: Rationality is up and emotion is down. *Journal of Consumer Research, 42,* 632-651.

Clark, H.H. (1973). Space, time, semantics, and the child. In T.E. Moore (Ed.), *Cognitive development and the acquisition of language* (27–63). New York: Academic Press.

Clark-Carter, D. (2004). *Quantitative psychological research: A student's handbook.* New York: Psychology Press.

Cohen, N., Moyal, N., & Henik, A. (2015). Executive control suppresses pupillary responses to aversive stimuli. *Biological Psychology, 112,* 1–11.

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General, 122,* 371–396.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56,* 5-18.

Dovidio, J., Kawakami, K., Johnson, C., Johnson, B. & Howard, A. (1997). On the nature of prejudice: Automatic and controlled processes. *Journal of Experimental Social Psychology, 33,* 510–540.

Eagly, A. H., Makhijani, M .G., & Klonsky, B.G. (1992). Gender and the evaluation of leaders: A meta-analysis. *Psychological Bulletin, 111,* 3-22.

Eagly, A.H., & Steffens, V. (1984). Gender stereotypes stem from the distribution of women and men into social roles. *Journal of Personality and Social Psychology, 46,* 735–754.

Ellemers, N. (2018). Gender stereotypes. *Annual Review of Psychology, 69,* 275-298.

Fairclough, S. H., & Mulder, L. J. M. (2011). Psychophysiological processes of mental effort investment. In R. A. Wright & G. H. E. Gendolla (Eds.), *How motivation affects cardiovascular response: Mechanisms and applications* (pp. 61–76). Washington, DC: American Psychological Association.

Fauconnier, G., & Turner, M. (1998). Conceptual integration networks. *Cognitive Science, 22,* 133-187.

Fazio, R. H. (1990). A practical guide to the use of response latency in social psychological research. In C. Hendrick & M. S. Clark (Eds.), *Research methods in personality and social psychology.* London: Sage.

Fiske, A. P. (2004). Four modes of constituting relationships: Consubstantial assimilation; space, magnitude, time, and force; concrete procedures; abstract symbolism. In N. Haslam (Ed.), *Relational models theory: A contemporary overview* (pp. 61–146). Mahwah, NJ: Erlbaum.

Fiske, S. T., & Lin, M. H., & Neuberg, S. L. (1999). The continuum model: Ten years later. In S. Chaiken & Y. Trope (Eds.), *Dual process theories in social psychology* (pp. 231-254). New York: Guilford Press.

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology, 82*, 878-902.

Fiske, S. T., & Taylor, S, E. (2010). *Social cognition: From brains to culture.* New York: McGraw Hill.

Fiske, S. T., & Lin, M. H., & Neuberg, S. L. (1999). The continuum model: Ten years later. In S. Chaiken & Y. Trope (Eds.), *Dual process theories in social psychology* (pp. 231-254). New York: Guilford Press.

Fried, R., Schettlinger, K., & Borowski, M. (2014). robfilter: Robust Time Series Filters. [Computer software]. Retrieved from https://CRAN.R-project.org/package=robfilter

Ford, T. E., & Stangor, C. (1992). The role of diagnosticity in stereotype formation: Perceiving group means and variances. *Journal of Personality and Social Psychology, 63*, 356-367

Galinsky, A. D., Gruenfeld, D. H., & Magee, J. C. (2003). From power to action. *Journal of Personality and Social Psychology, 85*, 453-466.

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*, 300-312.

Geva, R., Zivan, M., Warsha, A., & Olchik, D. (2013). Alerting, orienting or executive attention networks: Differential patters of pupil dilations. *Frontiers in Behavioral Neuroscience, 7*, 145.

Gevers, W., Caessens, B., & Fias, W. (2005). Towards a common processing architecture underlying Simon and SNARC effects. *European Journal of Cognitive Psychology, 17,* 659–673.

Giessner, S.R., & Schubert, T.W. (2007). High in the hierarchy: How vertical location and judgments of leaders' power are interrelated. *Organizational Behavior and Human Decision Processes, 104,* 30-44.

Gilbert, D. T., & Hixon, J. G. (1991). The trouble of thinking: Activation and application of stereotypic beliefs. *Journal of Personality and Social Psychology, 60*, 509–517.

Glick, P., & Fiske, S. T. (1996). The ambivalent sexism inventory: Differentiating hostile and benevolent sexism. *Journal of Personality and Social Psychology, 70,* 491– 512.

Glick, P., & Fiske, S. T. (2001). An ambivalent alliance. Hostile and benevolent sexism as complementary justifications for gender inequality. *American Psychologist, 56*, 109-118.

Greenwald, A.G., & Banaji, M.R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review, 102(2),* 4-27.

Greenwald, A.G., McGhee, D.E., & Schwartz, J.L.K. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology, 74,* 1464-1480.

Grunspan, D. Z, Eddy, S. L., Brownell, S. E., Wiggins, B. L., Crowe, A. J., Goodreau, S. M. (2016). Males under-estimate academic performance of their female peers in undergraduate biology classrooms. *PLoS ONE 11*, e0148405.

Guinote, A., & Phillips, A. (2010). Power can increase stereotyping: Evidence from managers and subordinates in the hotel industry. *Social Psychology, 41*, 3-9.

Guinote, A., Willis, G. B., Martellotta, C. (2010). Social power increases implicit prejudice external link. *Journal of Experimental Social Psychology, 46*, 299-307.

Guinote, A. (2017). How power affects people: Activating, wanting and goal seeking. *Annual Review of Psychology, 68*, 353–381.

Haddock, G., & Zanna, M. P. (1994). Preferring “housewives” to “feminists”: Categorization and the favorability of attitudes toward women. *Psychology of Women Quarterly, 18*, 25–52.

Hall, N. R. & Crisp, R. J. (2005). Considering multiple criteria for social categorization can reduce intergroup bias. *Personality and Social Psychology Bulletin, 31*, 1435–1444.

Hamilton, H. W., & Deese, J. (1971). Does linguistic marking have a psychological correlate? *Journal of Verbal Learning and Verbal Behavior, 10*, 707–714.

Hegarty, M. (2004). Mechanical reasoning by mental simulation. *Trends in Cognitive Sciences, 8*, 280–285.

Hess, E. H. & Polt, J. M. (1964). Pupil size in relation to mental activity during simple problem solving. *Science*, *140*, 1190-1192.

Hofmann, W., De Houwer, J., Perugini, M., Baeyens, F., & Crombez, G. (2010). Evaluative conditioning in humans: A meta-analysis. *Psychological Bulletin, 136*, 390-421.

Howitt, D., & Cramer, D. (2014). *Introduction to statistics in psychology*, 6th Edition. Harlow: Pearson Education.

Hyde, J. (2014). Gender similarities and differences. *Annual Review of Psychology*, *65*, 373–98.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*, 434-446.

Jamrozik, A., McQuire, *M*., Cardillo, E. R., & Chatterjee, A. (2016). Metaphor: Bridging embodiment to abstraction. *Psychonomic Bulletin and Review, 23,* 1080-1089.

Jarvis, B.G. (2012). DirectRT (Version v2012) [Computer Software]. New York, NY: Empirisoft Corporation.

Jarvis, B.G. (2012). MediaLab (Version v2012) [Computer Software]. New York, NY: Empirisoft Corporation.

Judd, C. M., Westfall, J., & Kenny D. A. (2012). Treating stimuli as a random factor insocial psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology, 103,* 54-69.

Judge, T. A., & Cable, D. M. (2004). The effect of physical height on workplace success and income: Preliminary test of a theoretical model. *Basic and Applied Social Psychology*, *89*, 428–441.

Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science, 154*, 1583-1585.

Katz, I., & Hass, R. G. (1986). Racial ambivalence and value conflict: Correlational and priming studies of dual cognitive structures. *Journal of Personality and Social Psychology, 55*, 893–905.

Kawakami, K., Dion, K. L., & Dovidio, J. F. (1998). Racial prejudice and stereotype activation. *Personality and Social Psychology Bulletin*, *24*, 407-416.

Kawakami, K., Dovidio, J. F., Moll, J., Hermsen, S., & Russin, A. (2000). Just say no (to stereotyping): Effects of training in the negation of stereotype associations on stereotype activation. *Journal of Personality and Social Psychology, 78*, 871–888.

Kloosterman, N. A., Meindertsma, T., van Loon, A. M., Lamme, V. A. F., Bonneh, Y. S., & Donner, T. H. (2015). Pupil size tracks perceptual content and surprise. *European Journal of Neuroscience, 41*, 1068-1078.

Kunda, Z., & Spencer, S. J. (2003). When do stereotypes come to mind and when do they color judgment? A goal-based theoretical framework for stereotype activation and application. *Psychological Bulletin, 129*, 522-544.

Laeng, B., Ørbo, M., Holmlund, T., & Miozzo, M. (2011). Pupillary Stroop effects. *Cognitive Processing, 12*, 13–21.

Lakens, D., & Etz, A. J. (2017). Too true to be bad: When sets of studies with significant and nonsignificant findings are probably true. *Social Psychological and Personality Science, 8*, 875-881.

Lakoff, G., & Johnson, *M*. (1999). *Philosophy in the flesh: The embodied mind and its challenge to western thought.* New York: Basic Books.

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D.H.J., Hawk, S. T., & van Knippenberg, A. (2010). Presentation and validation of the Radboud Faces Database. *Cognition & Emotion, 24,* 1377-1388.

Lebois, A.M, Wilson-Mendenhall, C.D., & Barsalou, L.W. (2015). Are automatic conceptual cores the gold standard of semantic processing? The context-dependence of spatial meaning in grounded congruency effects. *Cognitive Science, 39,* 1764-1801.

Lepore, L., & Brown, R. (1997). Category and stereotype activation: Is prejudice inevitable? *Journal of Personality and Social Psychology, 72*, 275-287.

Leslie, S. J, Cimpian A., Meyer, M., & Freeland, E. (2015). Expectations of brilliance underlie gender distributions across academic disciplines. *Science, 347*, 262–65.

Lewin, K. (1941). Regression, retrogression and development. *University of Iowa Studies of Child Welfare, 18*, 1–43.

Lin, H., Saunders, B., Hutcherson, C. A., & Inzlicht, M. (2018). Midfrontal theta and pupil dilation parametrically track subjective conflict (but also surprise) during intertemporal choice. *NeuroImage, 172,* 838-852.

Logel, C., Walton, G. M., Spencer, S. J., Iserman, E., von Hippel, W., & Bell, A. (2009). Interacting with sexist men triggers social identity threat among female engineers. *Journal of Personality and Social Psychology, 96,* 1089-1103.

Lundqvist, D., Flykt, A., & Öhman, A. (1998). *The Karolinska Directed Emotional Faces - KDEF, Department of Clinical Neuroscience*. Psychology section, Karolinska Institutet, ISBN 91-630-7164-9.

Maass, A., Suitner, C., Favaretto, X., & Cignacchi, M. Groups in space: Stereotypes and the spatial agency bias. *Journal of Experimental Social Psychology, 45*, 496-504.

MacNell, L., Driscoll, A., Hunt, A. N. (2015). What's in a name: Exposing gender bias in student ratings of teaching. *Innovative Higher Education, 40,* 291–303.

Macrae, C. N., & Bodenhausen, G. V. (2001). Social cognition: Categorical person perception. *British Journal of Psychology, 92*, 239-255

Macrae, C. N., Bodenhausen, G. V., Milne, A. B., Thorn, T. M. J., & Castelli, L. (1997). On the activation of social stereotypes. The moderating role of processing objectives. *Journal of Experimental Social Psychology, 33,* 471-489.

Mandler, J.M. (1992). How to build a baby: Conceptual primitives. *Psychological Review, 99*, 587–604.

Mason, C. M. (2006). Exploring the processes underlying within-group homogeneity. *Small Group Research, 37,* 233-270.

MacInnis, D. J. & Price, L. L. (1987). The role of imagery in information processing: Review and extensions. *Journal of Consumer Research, 13,* 473-491.

Mathôt, S., Fabius, J., Van Heusden, E., & Van der Stigchel, S. (2018). Safe and sensible preprocessing and baseline correction of pupil-size data. *Behavior Research Methods*, *50*, 94-106.

Medin, D. L. (1989). Concepts and conceptual structure. *American Psychologist, 44,* 1469-1481.

Meier, B. P., Hauser, D. J., Robinson, M.D., Friesen, C. K., Schjeldahl, K. (2007). What's "up" with God? Vertical space as a representation of the divine. *Journal of Personality and Social Psychology, 93(5),* 699-710.

Meier, B.P., & Robinson, M.D. (2004). Why the sunny side is up. *Psychological Science, 15*, 243-247.

Miller, S. S., & Saucier, D. A. (2018). Individual differences in the propensity to make attributions to prejudice. *Group Processes & Intergroup Relations, 21,* 280-301.

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*, 140-151.

Müller, F., & Rothermund, K. (2004). What does it take to activate stereotypes? Simple primes don’t seem enough. A replication of stereotype activation (Banaji & Hardin, 1996; Blair & Banaji, 1996). *Social Psychology, 45,* 187-193.

Niedenthal, P. M., Barsalou, L. W., Winkielman, P., Krauth-Gruber, S., Ric, F. (2005). Embodiment in attitudes, social perception, and emotion. *Personality and Social Psychology Review, 9*, 184-211.

Nisbett, R., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment.* Englewood Cliffs, N.J.: Prentice-Hall.

Olson, M. A., & Fazio, R. H. (2001). Implicit attitude formation though classical conditioning. *Psychological Science, 12,* 413-117.

Olson, M. A., & Fazio, R. H. (2006). Reducing automatically activated racial prejudice through implicit evaluative conditioning. *Personality and Social Psychology Bulletin*, *32,* 421-433.

Operario, D. & Fiske, S. (2008). Stereotypes: Content, structures, processes, and context. In R. Brown & S. Gaertner (Eds.), *Blackwell Handbook of Social Psychology: Intergroup Processes*, 22-44.

Payne, B. K., Vuletich, H. A., & Lunberg, K. B. (2017). The bias of crowds: How implicit bias bridges personal and systemic prejudice. *Psychological Inquiry, 28*, 233-248.

Plant, E.A., Hyde, J.S., Keltner, D., & Devine, P.G. (2006). The gender stereotyping of emotions. *Psychology of Women Quarterly*, *23,* 81-92.

Prehn, K., Heekeren, H. R., Blasek, K., Lapschies, K., Mews, I., & van der Meer, E. (2008). Neuroticism influences pupillary responses during an emotional interference task. *International Journal of Psychophysiology, 70,* 40-49.

Preuschoff, K., Hart, B. M., & Einhäuser, W. (2011). Pupil dilation signals surprise: Evidence for noradrenaline's role in decision making. *Frontiers in Neuroscience, 5*, 115.

Proctor, R. W., & Cho, Y. S. (2006). Polarity correspondence: A general principle for performance of speeded binary classification tasks. *Psychological Bulletin, 132,* 416 – 442.

Proctor, R. W., & Xiong, A. (2015). Polarity correspondence as a general compatibility principle. *Current Directions in Psychological Science, 24,* 446-451.

Proulx, T., Inzlicht, M., & Harmon-Jones, E. (2012). Understanding all inconsistency compensation as a palliative response to violated expectations. *Trends in Cognitive Sciences*, *16,* 285–291.

Proulx, T., Sleegers, W. and Tritt, S. M. 2017. The expectancy bias: Expectancy-violating faces evoke earlier pupillary dilation than neutral or negative faces. *Journal of Experimental Social Psychology*, *70,* 69-79.

Psychology Software Tools, Inc. [E-Prime 2.0]. (2016). Retrieved from https://www.pstnet.com.

R Core Team (2016). R: A language and environment for statistical computing [Computer software]. Retrieved from https://www.R-project.org.

Ratcliff, N. J., Vescio, T. K., (2013). Benevolently bowing out: the influence of self-construals and leadership performance on the wilful relinquishing of power. *Journal of Experimental Social Psychology*, *49,* 978–83.

Reinhard, G., & Lachnit, H. (2002). The effect of stimulus probability on pupillary response as an indicator of cognitive processing in human learning and categorization. *Biological Psychology, 60,* 199–215.

Rocklage, M.D., & Fazio, R. H. (2015). The Evaluative Lexicon: Adjective use as a means of assessing and distinguishing attitude valence, extremity, and emotionality. *Journal of Experimental Social Psychology, 56*, 214-227.

Roets, A., van Hiel, & Dhont (2012). Is sexism a gender issue? A motivated social cognition perspective on men's and women's sexist attitudes toward own and other gender. *European Journal of Personality, 26,* 350-359.

Rudman, L.A., & Kilianski, S.E. (2000). Implicit and explicit attitudes toward female authority. *Personality and Social Psychology Bulletin*, *26,* 1315-1328.

Rudman, L.A., Moss-Racusin, C.A., Phelan, J.E., & Nauts, S. (2012). Status incongruity and backlash effects: Defending the gender hierarchy motivates prejudice against female leaders. *Journal of Experimental Social Psychology, 48,* 165-179.

Rudman, L.A., & Phelan, J.E. (2010). The effects of priming gender roles on women's implicit gender beliefs and career aspirations. *Social Psychology, 41,* 192-202.

Rule, N. O., Adams Jr, R. B., Ambady, N., & Freeman, J. B. (2012). Perceptions of dominance following glimpses of faces and bodies. *Perception, 41*, 687-706.

Sani, F., Bennett, M., Mullally, S., & MacPherson, J. (2003). On the assumption of fixity in children's stereotypes: A reappraisal. *British Journal of Developmental Psychology, 99*, 113–124.

Schneider, I.K., Rutjens, B.T. & Jostmann, N.B. & Lakens, D. (2011). Weighty matters. Importance literally feels heavy. *Social Psychological and Personality Science, 2,* 474 – 478.

Schubert, T. (2005). Your highness: Vertical positions as perceptual symbols of power. *Journal of Personality and Social Psychology, 89,* 1-21.

Shapiro, L. (2011). *Embodied cognition.* New York: Routledge.

Shaver, P., Pierson, & Lang (1974). Converging evidence for the functional significance of imagery in problem solving. *Cognition, 3*, 359-375.

Sherman, J. W. (1996). Development and mental representation of stereotypes. *Journal of Personality and Social Psychology, 70,* 1126-1141.

Sinclair, L., & Kunda, Z. (1999). Reactions to a Black professional: Motivated inhibition and activation of conflicting stereotypes. *Journal of Personality and Social Psychology, 77,* 885–904.

Sleegers, W.W.A., Proulx, T., & van Beest, I. (2015). Extremism reduces conflict arousal and increases values affirmation in response to meaning violations. *Biological Psychology, 108,* 126–131.

Slepian, M.L, Weisbuch, M., Rule, N.O., & Ambady, N. (2011). Tough and tender: Embodied categorization of gender. *Psychological Science, 22(1),* 26-28.

Smith, L. & Gasser, M. (2006). The development of embodied cognition: Six lessons from babies. *Artificial Life, 11,* 13-29.

Smallwood, J., Brown, K.S., Tipper, C., Giesbrecht, B., Franklin, M.S., Mrazek, M.D., et al. (2011). Pupillometric evidence for the decoupling of attention from perceptual input during offline thought. *PLoS ONE 6,* e18298.

Spatola, N., Santiago, J., Beffara, B., Mermillod, M., Ferrand, L., & Ouellet, M. (2018). When the sad past is left: The mental metaphors between time, valence, and space. *Frontiers in Psychology, 9,* 1019.

Sriram, N., & Greenwald, A.G. (2009). The Brief Implicit Association Test. *Experimental Psychology, 56*, 283-294.

Stewart, L. H., Ajina, S., Getov, S., Bahrami, B., Todorov, A., & Rees, G. (2012). Unconscious evaluation of faces on social dimensions. *Journal of Experimental Psychology: General, 141*, 715

Suitner, C. & Maass, A. (2016). Spatial agency bias: Representing people in space. In: M. Zanna, & J. Olson (Eds.), *Advances in Experimental Social Psychology, 53*, 245-301.

Taylor, J. E., Lam, T., Chasteen, A., Pratt, J. (2015). Bow your head in shame, or, hold Your head up with pride: Semantic processing of self-esteem concepts orients attention vertically. *PLoS ONE, 10,* e0137704.

Tobii Group. (2015, April 27). Retrieved May 14, 2018, from https://www.tobii.com.

Trevino, L. J., Gomez-Mejia, L. R, Balkin, D. B, & Mixon, F. G. (2015). Meritocracies or masculinities? The differential allocation of named professorships by gender in the academy. *Journal of Management, 44,* 972-1000.

van der Wel, P. & van Steenbergen, H. (2018). Pupil dilation as an index of effort in cognitive control tasks: A review. *Psychonomic Bulletin & Review,* 1-11.

van Gerven, P.W.M., Paas, F., van Merriënboer, J. J. G., & Schmidt, H. G. (2004). Memory load and the cognitive pupillary response in aging. *Psychophysiology, 41*, 167–174.

van Steenbergen, H., & Band, G. P. H. (2013). Pupil dilation in the Simon task as a marker of conflict processing. *Frontiers in Human Neuroscience, 7,* 215.

von Hecker, U., Klauer, C.K. & Sankaran, S. (2013). Embodiment of social status: Verticality effects in multi-level rank-orders. *Social Cognition, 31,* 374-389.

von Hecker, U., Klauer, K.C., Wolf, L., & Fazilat-Pour, *M*. (2016). Spatial processes in linear ordering. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42,* 1003-1033.

Weeks, D. J., & Proctor, R. W. (1990). Salient-features coding in the translation between orthogonal stimulus-response dimensions. *Journal of Experimental Psychology: General, 119,* 335-366.

Wilder, D. A. (1981) Perceiving persons as a group: Categorization and ingroup relations. In D. L. Hamilton (Ed.), *Cognitive processes in stereotyping and intergroup behavior* (pp. 213-257). Hillsdale, NJ: Erlbaum.

Williams, L.E., Huang, J. Y., & Bargh, J. A. (2009). The scaffolded mind: Higher mental processes are grounded in early experience of the physical world. *European Journal of Social Psychology, 39,* 1257-267.

Wood, W., & Eagly, A.H. (2002). A cross-cultural analysis of the behaviour of women and men: Implications for the origins of sex differences. *Psychological Bulletin, 128*, 699-727.

Yap, A. J., Mason, M. F., & Ames, D. R (2013). The powerful size others down: The link between power and estimates of others' size. *Journal of Experimental Social Psychology*, *49*, 591-594.

Zanolie, K., van Dantzig, S., Boot, I., Wijnenc, J., Schubert, T., Giessner, S., Pecher, P. (2012). Mighty metaphors: Behavioural and ERP evidence that power shifts attention on a vertical dimension. *Brain and Cognition, 78,* 50-58.

Zarzeczna, N., von Hecker, U., & Haddock, G. (2018a). *Males on top: Cognitive blending involves spatial simulations dependent on stereotypic thinking.* Manuscript under review. Manuscript in preparation.

Zarzeczna, N., von Hecker, U., & Haddock, G. (2018b). *Males on top prime perception of greater sizes: Spatial location and size are associated with mental representations of power.* Manuscript under review.

Zink, C. F., Tong, Y., Chen, Q., Bassett, D. S., Stein, J. L., & Meyer-Lindenberg, A. (2008). Know your place: Neural processing of social hierarchy in humans. *Neuron, 58*, 273-283.

Appendixes

Appendix 1

Procedure of modelling effects

In order to establish the random structure of the final linear mixed model for each study, I first estimated a minimal model, in which the intercepts varied across *participants.* Subsequently, I estimated three other models that had a similar structure to the minimal model, except that in each model, I also introduced one random slope to test whether random effects improved the fit of my data in the final model. That is, in Model 1, I introduced *gender Choice* (males and females as powerful and powerless) as a random slope, in Model 2, I included *trial type* (males-top or females-top). After estimating the models, I compared the fit of each model that included a random slope with the minimal model by using Chi-square difference statistic $\Delta\chi 2$. If a model with a random slope provided a better fit for my data (i.e., the loglik ratio was significantly smaller than the ratio of the minimal model), the given random slope was then kept in the final model. The comparisons between minimal and random slope models as well as the structure of final models for each study are presented in tables below.

Study 1.

Model	df	AIC	BIC	loglik	deviance	$\Delta\chi 2$	Δdf	*p*
Minimal Model	10	42157	42217	-21069	42137			
Model 1 Gender choice	12	42156	42228	-21066	42132	5.54	2	.06
Model 2 Trial type	12	42147	42219	-21062	42123	14.24	2	.001*

The final model included *trial type* as a random slope.

Study 2.

Model	df	AIC	BIC	loglik	deviance	Δ$\chi 2$	Δdf	*p*
Minimal Model	10	40866	40925	-20423	40846			
Model 1 Gender choice	12	40868	40939	-20422	40844	1.54	2	.46
Model 2 Trial type	12	40869	40941	-20423	40845	.50	2	.78

The final model did not include any random slopes.

Study 3.

Model	df	AIC	BIC	loglik	deviance	Δ$\chi 2$	Δdf	*p*
Minimal Model	10	27885	27941	-13933	27865			
Model 1 Gender choice	12	27883	27950	-13930	27859	6.01	2	.05*
Model 2 Trial type	12	28048	27889	-27956	27865	.01	2	.99

The final model included *gender choice* as a random slope.

Study 4.

Model	df	AIC	BIC	loglik	deviance	Δ$\chi 2$	Δdf	*p*
Minimal Model	18	60161	60275	-30062	60125			
Model 1 Gender choice	20	60153	60280	-30057	60113	11.65	2	.01*
Model 2 Trial type	20	60164	60290	-30062	60124	1.0	2	.61
Model 3 Prime	20	60162	63000	-30061	60122	2.60	2	.27

The final model included *gender choice* as random slope.

Study 5.

Model	df	AIC	BIC	loglik	deviance	Δ$\chi 2$	Δdf	*p*
Minimal Model	10	33065	33122	-16522	33045			
Model 1 Gender choice	12	33023	33092	-16500	32999	45.80	2	.001*
Model 2 Trial type	12	33069	33138	-16522	33045	.16	2	.93

The final model included *gender choice* as a random slope.

Study 6.

Model	df	AIC	BIC	loglik	deviance	Δ$\chi 2$	Δdf	*p*
Minimal Model	10	32471	32529	-16226	32451			
Model 1 Gender choice	12	32471	32539	-16223	32447	4.64	2	.99
Model 2 Trial type	12	32473	32542	-16225	32449	3.01	2	.37

The final model did not include any random slopes.

Studies 1-6 – integrative analysis

Model	df	AIC	BIC	loglik	deviance	Δ$\chi 2$	Δdf	*p*
Minimal Model	50	238639	239025	-119270	238539			
Model 1 Gender choice	52	238599	238999	-119247	238495	44.90	2	.001*
Model 2 Trial type	52	239042	239042	-119268	238537	2.70	2	.26

The final model included *gender choice* as a random slope.

Study 7.

(a) Schubert block (overall analysis)

Model	df	AIC	BIC	loglik	deviance	$\Delta\chi 2$	Δdf	*p*
Minimal Model	6	184	214	-85	172			
Model 1 Position	8	184	225	-84	168	3.16	2	.21
Model 2 Response period	8	155	155	-49	98	73.0	2	.001*

The final model included *response period* as a random slope.

(b) Gender block (overall analysis)

Model	df	AIC	BIC	loglik	deviance	$\Delta\chi 2$	Δdf	*p*
Minimal Model	6	115	166	-48	95			
Model 1 Gender choice	8	105	166	-40	81	14.41	2	.001*
Model 2 Trial type	8	118	179	-47	94	1.32	2	.52
Model 3 Response period	8	35	97	-6	11	84.11	2	.001*

Study 8.

Model	df	AIC	BIC	loglik	deviance	$\Delta\chi 2$	Δdf	*p*
Minimal Model	10	69519	69584	-34750	69499			
Model 1 Gender block	12	69381	69459	-34679	69357	141.94	2	.001*
Model 2 Position choice	12	69518	69596	-34747	69494	5.07	2	.08

The final model included *gender block* as a random slope.

Study 9.

Model	df	AIC	BIC	loglik	deviance	Δ*χ2*	Δdf	*p*
Minimal Model	10	69242	69307	-34611	69222			
Model 1 Gender block	12	69117	69195	-34546	69093	129.67	2	.001*
Model 2 Position choice	12	69246	69324	-34611	69222	.27	2	.87

The final model included *gender block* as a random slope.

Study 10.

Model	df	AIC	BIC	loglik	deviance	Δ*χ2*	Δdf	*p*
Minimal Model	10	50986	51050	-25483	50966			
Model 1 Gender	12	50983	51059	-25479	50959	7.62	2	.02 *
Model 2 Vertical Position	12	50984	51060	-25480	50960	6.28	2	.04 *

The final model included *gender* and *vertical position* as random slopes.

Study 12.

Model	df	AIC	BIC	loglik	deviance	Δ*χ2*	Δdf	*p*
Minimal Model	10	24449	24503	-12214	24429			
Model 1 Gender choice	12	24448	24512	-12212	24424	5.16	2	.08
Model 2 Trial type	12	24452	24517	-12214	24428	.17	2	.92

The final model did not include any random slopes.

Study 14.

Model	df	AIC	BIC	loglik	deviance	$\Delta\chi 2$	Δdf	*p*
Minimal Model	10	34787	34845	-17383	34767			
Model 1 Gender choice	12	34777	34847	-17377	34753	13.58	2	.001*
Model 2 Trial type	12	34791	34860	-17383	34767	.01	2	.99

The final model included gender choice as a random slope.

Appendix 2

Method of individual studies (1 – 6)

Studies 1 and 2

Materials

Spatial task. I selected ten popular British (Study 1) and Polish (Study 2) names (five female and five male).[33] Each name was randomly paired with a name of the opposite gender (e.g., Oliver – Emily). The matched pairs were then presented on the computer screen in white letters on a black screen (font size 15 – 21 points).

Design

Spatial task. The same design was used in both studies. Task condition (find powerful versus powerless) was manipulated between-participants while the trial type (males-top versus females-top) was manipulated within-participants. There were four blocks of trials. Each block included a presentation of five pairs shown twice (10 trials in total). In half of the trials, a male name (e.g., Oliver) was displayed at the top of the screen with a female name (e.g., Emily) at the bottom. Both names were centred

[33] See the most popular names: Popular British baby names: Year by year. (n.d.). Retrieved February 9, 2015, from http://www.babycentre.- co.uk/popular- baby-names (Study 1); Academy of childbirth. (n.d.). Retrieved May 2, 2016, from: http://akademiaporodu.pl/top-news/najpopularniejsze-imiona-w-2015-2016-mapy-ranking (Study 2).

and there was a 23cm vertical distance between the names. In the other trials, this display was reversed. There were 40 trials in total. The order of trials within each block was randomized.

Social Status - Gender IAT. I adapted the IAT of implicit attitudes towards high or low status males and females from research by Rudman and Kilianski (2000). Within a single block participants were asked to categorise items that belonged to either high status/males or low status/females, or vice versa in another block. The order of the blocks was counterbalanced across participants.

Procedure

The studies were presented using DirectRT (Jarvis, 2012). Participants sat approximately 70cm away from the computer screen. After reading the instructions, I verbally explained the concept of power to participants by adapting Galinsky, et al. (2003) definition of social power. I explained the definition verbally to ensure that participants understood the definition of social power and to give them an opportunity to ask questions about it. Subsequently, participants completed the spatial task by pressing A or L on the computer keyboard to indicate the powerful/powerless person on the screen. In Study 2, the procedure was the same, but all instructions and stimuli were translated into Polish. After the spatial task, all participants completed the IAT.[34] Each experiment lasted approximately 15 minutes.

Study 3

Method

Materials

Spatial Task. The selected professions (derived from a pilot study) were assigned to male and female names (adapted from Study 1; e.g., Oliver-Professor, Sophie-Professor). Each pair (profession and name) was assigned to a corresponding

[34] Within all the reported studies, participants signed consent forms at the beginning of experiments and were debriefed at the end.

pair including a name of member of the opposite gender. I created all combinations of professions and genders (e.g., male scientist - female professor; female professor - male scientist) within four sets of pairings. Participants were randomly assigned to receive one set of combinations.

Design

Spatial Task. I used the same mixed design as in Study 1 and 2 and presented participants with the same spatial task, except that in half the trials, presented in a random order, a male name and a profession were displayed at the top of the screen (e.g., Oliver-Professor) while a female name and a profession (e.g., Emily-Architect) at the bottom, whilst for the remaining half of the trials, this spatial orientation was reversed. To facilitate responses, I asked participants to use arrows (up and down). Using these arrows provides a congruent mapping of mental simulations onto motor responses (see Schubert, 2005).

Social Status - Gender IAT. I used the same IAT of attitudes towards high status males and females, as in Study 1, but I reduced the number of blocks from five to two (see Sriram & Greenwald, 2009). Participants were asked to keep in mind two target categories within two blocks (male and high status in one block or female and high status in another) and respond as fast as possible by pressing a designated key when they saw an item that belonged to the target category on the screen. When they saw distractors (female and low status or males and low status) they were supposed to press another key. The order of blocks was counterbalanced across participants.

Procedure

First, participants completed the spatial task and then the IAT. Both tasks were programmed as in Studies 1 and 2. The experiment took approximately 15 minutes.

Study 4

Method

Materials

Spatial Task. I displayed the same gender pairs as in Study 1, but before each trial participants were presented with a social status item (gender-neutral profession adapted from Study 3, i.e.: scientist, architect, doctor, professor, dentist) or neutral word (i.e., vegetables: carrot, potato, lettuce, broccoli, cabbage). The prime was manipulated within-subjects. After making decisions about powerful/powerless person in each trial, participants were asked to report whether the initially presented prime word belonged to the social status (by pressing the arrow pointing left) or vegetable category (by pressing the arrow pointing right).[35] In this way, the prime word was activated in participants' minds during their decision about the powerful/powerless person.

Social Status - Gender IAT. I adapted the IAT from Study 1.

Design

Spatial Task. I used the same mixed design as in Study 1 and 2, and presented participants with a similar spatial task, except that before half of the trials (males-top; females-bottom, n = 5) they were primed with a social status item, and before the other half (males-top; females-bottom, n = 5) they were shown a neutral word. The same was done for the trials where females were presented at the top and males at the bottom, so there were 20 trials in total within each block. The trials were randomised within each block within-participants; there were 4 blocks in total.

Social Status - Gender IAT. I used the same IAT of gender authorities as in Study 1.

[35] The responses were reversed for the other half of our participants: they pressed the arrow pointing right for social status items and the arrow pointing left for vegetables (participants were randomly assigned to one of the key arrangements).

Procedure

First, participants completed the spatial task and then the IAT. The tasks were programmed using DirectRT (Jarvis, 2012). The experiment took 15 minutes.

Studies 5 and 6

Method

Materials and design

Spatial task. I used the same stimuli as in Study 1. However, participants responded with the arrow up to indicate the powerful/powerless person at the top or the arrow pointing down to indicate the target at the bottom in Study 5. Study 6 was a replication of Study 1, so participants responded with horizontally arranged keys on the keyboard.

Rationality/emotionality and gender IAT. I asked participants to quickly and accurately sort items that belonged to the category of emotionality (feeling, mood, sentiment, sympathy) and females (e.g., Karen, Caitlin, Jasmine), as well as rationality (intelligent, logic, reason, thinking) and males (e.g., Brian, Kevin, Paul), and vice versa.

Rationality/emotionality and verticality IAT. Participants categorized items that belonged to the category of up (above, top, over, upper) and rationality and categories down (below, under, bottom, lower) and emotionality, or up/emotionality and down/rationality (the items for emotionality/rationality categories were the same as in the rationality/emotionality and gender IAT).

Both IATs were designed in the same standard way as the Social Status - Gender IAT used in Study 1.

Procedure

First, participants completed the spatial task, then the rationality/emotionality and gender IAT, which was followed by the IAT measuring the rationality/emotionality and verticality associations. I presented the tasks in the same order to all participants.

Appendix 3

Studies 1 and 3: Five matched pairs of names

Pair 1. Oliver; Emily

Pair 2. Jack; Lily

Pair 3. Thomas; Chloe

Pair 4. Jacob; Alice

Pair 5. James; Sophie

Study 2: Five matched pairs of names in Polish

Pair 1. Adam; Julia

Pair 2. Jakub; Emilia

Pair 3. Szymon; Alicja

Pair 4. Dawid; Marta

Pair 5. Piotr; Anna

Appendix 4

Study 3: Four sets of assigned professions and genders

Set 1.

Pair 1. Oliver Professor; Emily Architect

Pair 2. Jack Doctor; Lily Scientist

Pair 3. Thomas Architect; Chloe Dentist

Pair 4. Jacob Dentist; Alice Doctor

Pair 5. James Scientist; Sophie Professor

Set 2.

Pair 1. Thomas Architect; Sophie Professor

Pair 2. James Scientist; Alice Doctor

Pair 3. Jacob Dentist; Emily Architect

Pair 4. Jack Doctor; Chloe Dentist

Pair 5. Oliver Professor; Lily Scientist

Set 3.

Pair 1. Jacob Dentist; Lily Scientist

Pair 2. Thomas Architect; Alice Doctor

Pair 3. James Scientist; Emily Architect

Pair 4. Jack Doctor; Sophie Professor

Pair 5. Oliver Professor; Chloe Dentist

Set 4.

Pair 1. Jacob Dentist; Sophie Professor

Pair 2. Jack Doctor; Emily Architect

Pair 3. Oliver Professor; Alice Doctor

Pair 4. James Scientist; Chloe Dentist

Pair 5. Thomas Architect; Lily Scientist

Appendix 5

Studies 1 and 4: IAT categories and items

Female names: Karen, Caitlin, Jasmine, Charlotte, Ruby, Mary, Lauren, Kate, Zoe, Lara, Ann, Victoria, Bethany, Daisy, Sarah.

Male names: Brian, Kevin, Paul, Benjamin, Freddie, Joseph, Jake, Edward, Robert, Lewis, Toby, Liam, Patrick, Tommy, Arthur.

High status: Boss, supervisor, expert, leader, executive, authority.

Low status: Secretary, helper, aide, clerk, subordinate, assistant.

Study 2: IAT categories and items

Female names: Hanna, Paulina, Barbara, Justyna, Magdalena, Helena, Weronika, Klaudia, Dominika, Zofia, Zuzanna, Katarzyna, Aleksandra, Joanna, Oliwia.

Male names: Maciej, Jan, Filip, Wiktor, Gabriel, Marek, Konrad, Karol, Tomasz, Bartosz, Wojciech, Krzysztof, Patryk, Hubert, Adrian.

High status: Zwierzchnictwo, kierownictwo, nadzór, przywództwo, autorytet, dyrekcja.

Low status: Asysta, podporządkowanie, zależność, podrzędność, podwładność, służba.

Appendix 6

Results of individual studies (1 - 6)

Study 1

Spatial Task. Proportions. First, I conducted a chi-square test to test how often participants indicated gender across the task conditions. They were more likely to categorise males as powerful (68.7%) than females as powerful (31.3%), $\chi^2(1) = 272.67$, $p < .001$, odds ratio $_{\text{choosing males as powerful versus females}}$ = 1.80 (95%CI[1.67, 1.94]).

Spatial Task. Response Latencies. Subsequently, I estimated a linear mixed model to analyse the response latencies of participants' choices.[36] I found that the fixed effect of gender choice, $F(1, 2799) = 2.40$, $p = .12$, $dz = .03$, was not significant, as was the fixed effect of task, $F(1, 76) = .10$, $p = .75$, $dz = .06$. However, trial type was marginally significant, $F(1, 84) = 2.95$, $p = .09$, $dz = .16$, such that participants were marginally quicker to categorise powerful and powerless persons when they were presented with males at the top (*M* = 995ms, *SE* = 42.28, 95%CI[926, 1094]) in contrast to females at the top (*M* = 1025ms, *SE* = 47.09, 95%CI[942, 1130]).

In terms of two-way interactions, trial type by gender choice interaction was not significant, $F(1, 1393) = .40$, $p = .53$. Further, the gender choice by task interaction was significant, $F(1, 2799) = 9.05$, $p < .01$. Participants were significantly slower to categorise males as powerful (*M* = 1031ms, *SE* = 63.44, 95%CI[879, 1131]) versus females as powerful (*M* = 1022ms, *SE* = 64.96, 95%CI[941, 1198]), $t(2672) = -3.08$, $p < .01$, $dz = .34$. Also, they were quicker to categorise females as powerless (*M* = 966ms, *SE* = 64.96, 95%CI[941, 1198]) than males as powerless (*M* = 1033ms, *SE* = 64.96,

[36] See Appendix 1 – for the final structure of the model estimated for each individual study.

95%CI[941, 1198]), however, the difference was not significant, $t(2849) = -1.08$, $p = .28$, $dz = .12$.[37]

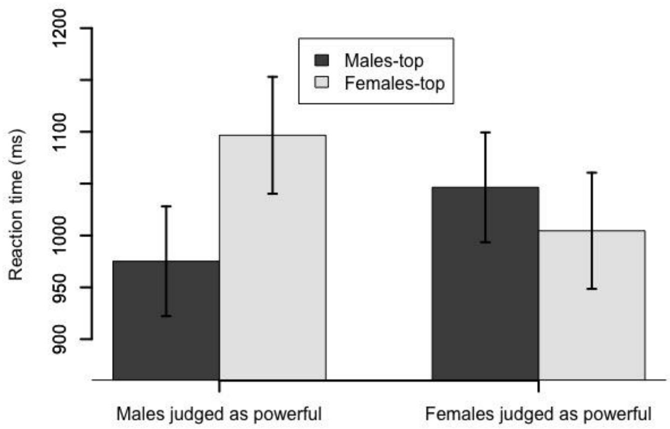
Next, the trial type by task interaction was marginally significant, $F(1, 84) = 3.27$, $p = .07$. Post-hoc comparisons revealed that participants were significantly quicker to categorise gender in the powerful condition on males-top trials (M = 994ms, SE = 60.70, 95%CI[890, 1132]), as opposed to females-top trials (M = 1063ms, SE = 67.46, 95%CI[929, 1198]), $t(93) = -2.41$, $p < .01$, $dz = .30$. The same difference was not significant in the powerless condition, (males-top: M = 996ms, SE = 58.87, 95%CI[892, 1127]; females-top: M = 987ms, SE = 65.73, 95%CI[877, 1139]), $t(82) = .07$, $p < .95$, $dz = .01$).

Inconsistent with my hypothesis, the three-way interaction between gender choice, trial type, and task was not significant, $F(1, 1393) = .59$, $p = .44$ (see Appendixes: Figure 1).

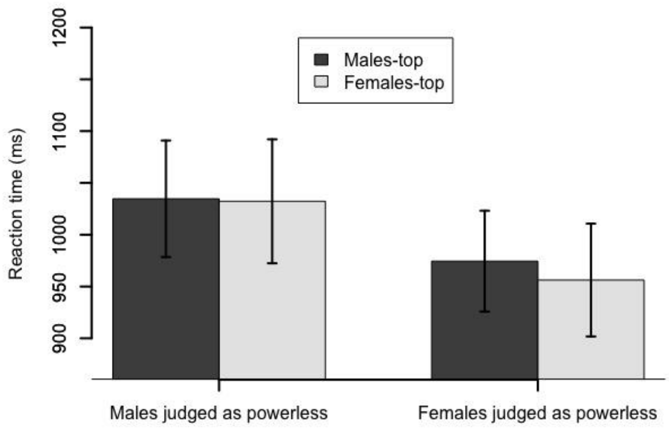
[37] The difference in actual means is larger in the powerless condition (females as powerless versus males as powerless) than in the powerful condition (males as powerful versus females as powerful), however, the linear mixed model estimations of mean differences (lsmeans) demonstrated that the estimated mean difference was larger in the latter.

Figure 1. Response latencies to (a) males and females judged as powerful, and (b) males and females judged as powerless according to the trial type (males-top/females-bottom or females-top/males-bottom). The error bars show -/+1 standard error.[38]

(a) Powerful condition



(b) Powerless condition



[38] Note that "Males-top" refers to trials where males were presented at the top and females at the bottom, whilst "Females-top" represents trials where females were presented at the top and males at the bottom.

Study 2

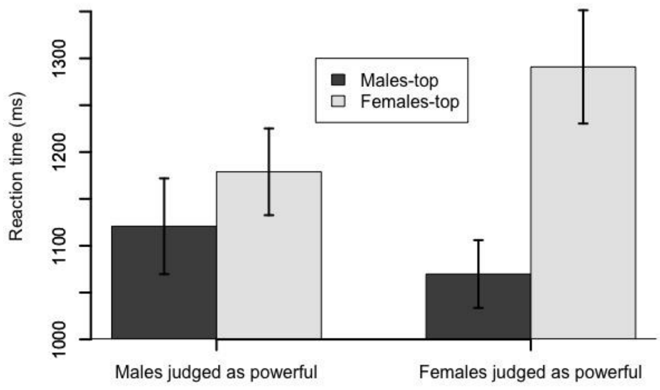
Spatial Task. Proportions. As in Study 1, participants had a tendency to indicate males as powerful (52.7%) as opposed to females (47.3%), $\chi^2(1) = 24.34$, $p < .001$, odds ratio $_{\text{choosing males as powerful versus females}}= 1.21$ (95%CI[1.12, 1.31]).

Spatial Task. Response Latencies. The results were not consistent with Study 1 and did not support my hypotheses. First, the fixed effects of gender choice $F(1, 2777) = 1.26$, $p = .26$, $dz = .01$, and task were not significant, $F(1,72) = .04$, $p = .85$, $dz = .02$. However, the fixed effect of trial type was significant, $F(1, 2627) = 4.22$, $p <0 .04$, dz = .17, suggesting that participants were significantly quicker to categorise gender as powerful/powerless on males-top trials, (M = 1116ms, SE = 40.02, 95%CI[1099, 1258]), as opposed to females-top trials, (M = 1232ms, SE = 39.29, 95%CI[1129, 1286]).

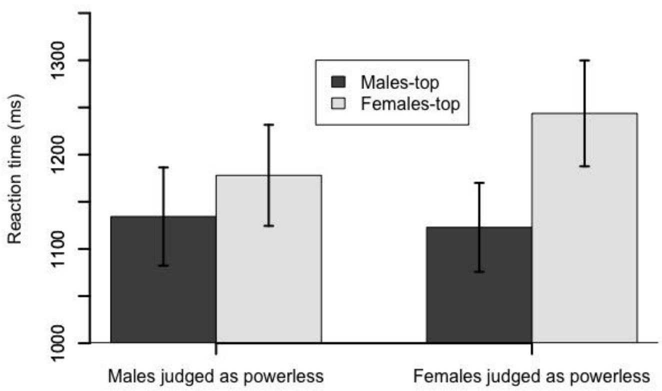
Two-way interactions (gender choice x task; gender choice by trial type; task by trial type) were not significant, $ps > .14$. The same was true for the three-way interaction (gender choice x trial type x task), $F(1, 2764) = .64$, $p = .42$ (see Appendixes: Figure 2).

Figure 2. Response latencies to (a) males and females judged as powerful, and (b) males and females judged as powerless according to the trial type (males-top/females-bottom or females-top/males-bottom). The error bars show -/+1 standard error.

(a) Powerful condition



(b) Powerless condition



Study 3

Spatial Task. Proportions. In contrast to the previous studies, in Study 3, participants were equally likely to categorise males (50.5%) and females (49.5%) as powerful, $\chi^2(1) = .27$, $p = .60$, odds ratio $_{\text{choosing males as powerful versus females}}$ = 1.02 (95%CI[.94, 1.20]).

Spatial Task. Response Latencies. The fixed effects of gender choice $F(1, 185) = .05$, $p = .83$, $dz = .01$, and trial type were not significant, $F(1, 1886) = 1.44$, $p = .23$, $dz = .06$. However, the fixed effect of task was significant, $F(1, 49) = 4.56$, $p < .04$, $dz = .39$, indicating faster response times in the powerful (*M* = 1128ms, *SE* = 35.32, 95%CI[1109, 1250]), versus powerless condition, (*M* = 1279ms, *SE* = 35.61, 95%CI[1151, 1293]).

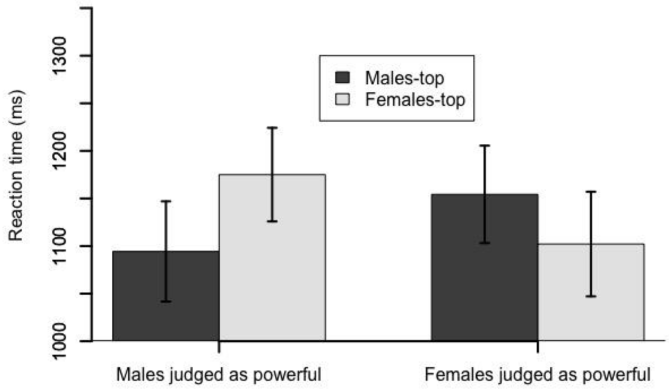
Further, the gender choice by trial type interaction was marginally significant, $F(1, 1889) = 3.47$, $p = .06$, such that participants were faster to categorise males as powerful and powerless when they were presented at the top (*M* = 1173ms, *SE* = 51.25, 95%CI[1176, 1382]) than bottom, (*M* = 1231ms, *SE* = 51.25, 95%CI[1176, 1382]), $t(1876) = -2.17$, $p < .03$, $dz = .44$). The same difference was not significant when participants categorised females at the top (*M* = 1194ms, *SE* = 54.40, 95%CI[1119, 1280]), and bottom(*M* = 1221ms, *SE* = 56.19, 95%CI[1128, 1290]), $t(1868) = .48$, $p = .63$, $dz = .05$). The other two-way interactions (gender choice by task and trial type by task) were not significant, *ps* > .37.

As expected, the three-way interaction (gender choice by trial type by task) was significant, $F(1, 1889) = 7.16$, $p < .01$ (see Appendixes: Figure 3). Post-hoc comparisons indicated that participants were quicker to categorise males as powerful at the top (*M* = 1094ms, *SE* = 55.76, 95%CI[981, 1179]), versus bottom, (*M* = 1175ms, *SE* = 59.09, 95%CI[1055, 1256]), $t(1870) = -2.77$, $p < .01$, $dz = .43$. The same pattern was true for females chosen as powerful, but the difference was marginally significant –

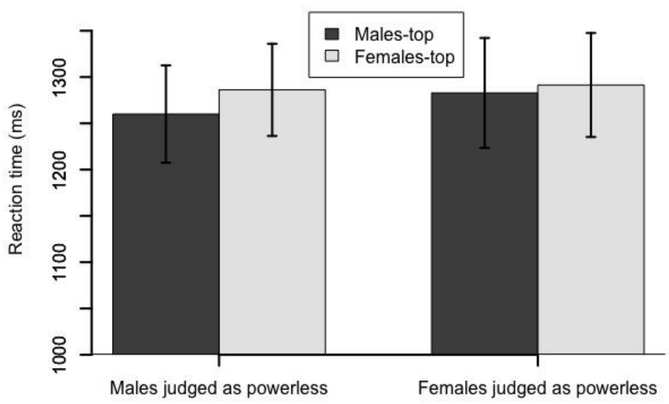
females were categorised marginally faster as powerful at the top (*M* = 1102ms, *SE* = 54.39, 95%CI[996, 1221]) versus bottom, (*M* = 1154ms, *SE* = 57.03, 95%CI[1045, 1273]), $t(1870) = -2.77$, $p < .01$, $dz = .30$. Other comparisons were not significant, *ps* > .25.

Figure 3. Response latencies to (a) males and females judged as powerful, and (b) males and females judged as powerless according to the trial type (males-top/females-bottom or females-top/males-bottom). The error bars show -/+1 standard error.[39]

(a) Powerful condition



(b) Powerless condition



[39] Note that "Males-top" refers to trials where males were presented at the top and females at the bottom, whilst "Females-top" represents trials where females were presented at the top and males at the bottom.

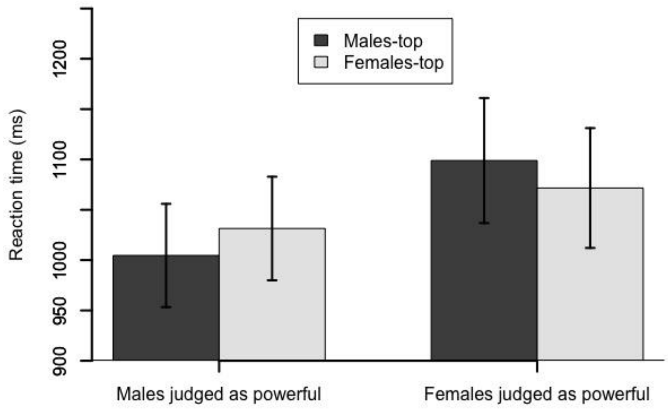
Study 4

Spatial Task. Proportions. In contrast to Study 3, in Study 4, I found that participants were more likely to pick males as powerful (64.9%) rather than females (35.1%), $\chi^2(1) = 170.59$, $p < .001$, odds ratio $_{\text{choosing males as powerful versus females}} = 1.44$ (95%CI[1.36, 1.53]).

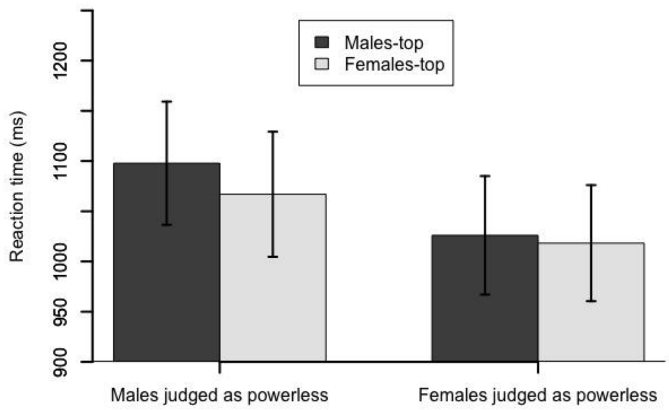
Spatial Task. Response Latencies. None of the fixed effects (gender choice, prime, task, and gender position) were significant, *ps* > .21. The same was true for two-way and three-way interactions, including the predicted interaction among gender choice, trial type, and task (see Figure 4), *ps* > .15. The predicted four-way interaction among gender choice, trial type, task, and prime was also not significant, $F(1, 4051) = 1.73$, $p = .19$. The only significant interaction was observed between gender choice and task, $F(1, 50) = 4.74$, $p < .03$. Participants were significantly faster to categorise males as powerful, (M = 1017ms, SE = 57.03, 95%CI[915, 1146]), in contrast to females (M = 1086ms, SE = 62.75, 95%CI[964, 1210]), $t(53) = -2.187$, $p < .02$, $dz = .35$, but the same difference was not significant in the case of categorising powerless individuals, $p > .51$ (see Appendixes: Figure 4)

Figure 4. Response latencies to (a) males and females judged as powerful, and (b) males and females judged as powerless according to the trial type (males-top/females-bottom or females-top/males-bottom). The error bars show -/+1 standard error.

(c) Powerful condition



(d) Powerless condition



Social Status - Gender IAT. Studies 1 – 4.

IATs: I found that participants were significantly faster at categorizing males and high status items (*M* = 712ms, *SE* = 12.61) compared to females and high status items (*M* = 756ms, *SE* = 16.02), $t(77) = 2.95$, $p < .01$, $dz = .30$.[40] The same pattern was observed in Study 2, $t(72) = -3.32$, $p < .01$, $dz = .40$. Again in Study 4, the pattern of results was the same, but the difference was marginally significant, $t(57) = -1.82$, $p = .07$, $dz = .27$. In Study 3, the difference between responses to males-high status versus females-high status, was not significant, $t(72) = -3.32$, $p < .01$, $dz = .07$.

Correlations: Subsequently, in each study, I calculated a facilitation score for each participant (by subtracting RTs from blocks in which participants responded to the category combination males/high status, from blocks in which they responded to females/high status; positive scores indicated a tendency to associate males with high status). Linking this with my RT data, I also computed two separate difference scores subtracting the RTs for choices of males at the top from the RTs for choices of males at the bottom and, vice versa, the RTs for choices of females at the bottom from the RTs for choices of females at the top, as it was done for the integrative analysis. I found that the IAT facilitation scores (i.e., the tendency to associate males with high status) did not correlate with the spatial simulation of males as powerful at the top: Study 1: $r(32) = .17$, $p = .36$; and of females as powerless at the bottom, $r(39) = -.05$, $p = .79$; Study 2: $r(35) = -.17$, $p = .35$ (males as powerful at the top), but there was a significant negative correlation between simulation of females as powerless at the bottom and the IAT facilitation score $r(35) = -.36$, $p < .02$. In Study 3, there were no correlations between the IAT facilitations core and simulations of males as powerful at the top, $r(24) = -.01$,

[40] All the IAT analyses in my studies were based on log-transformed data. However, I report mean reaction times using raw scores.

p = .96, and simulations of females as powerless at the bottom, *r*(25) = -.18, *p* = .40. The same was true for Study 4, (powerful-males-top: *r*(28) = .12, *p* = .54; powerless-females-bottom: *r*(28) = .20, *p* = .31).

Study 5: Results

Spatial Task. Proportions. As in the previous studies (except for Study 3), I found that participants picked males as powerful (70.3%) more often than females as powerful (29.7%), $\chi^2(1) = 313.07$, *p* < .001, *d* > 2.0, odds ratio $_{\text{choosing males as powerful versus females}}$ = 2.10 (95%CI[1.92, 2.30]).

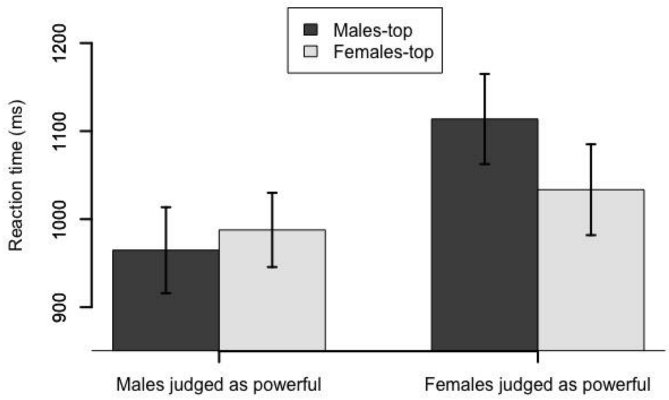
Spatial Task. Response latencies. The fixed effects of gender choice, trial type, and task were not significant, *ps* > .14. There were two marginally significant two-way interactions. First, the gender choice by trial type interaction, *F*(1, 2241) = 3.02, *p* = .08, indicated that participants were significantly quicker to categorise males when they were presented at the top, (*M* = 1009ms, *SE* = 33.03, 95%CI[965, 1098]), in contrast to females at the bottom, (*M* = 1084ms, *SE* = 41.40, 95%CI[1012, 1180]), *t*(89) = -2.12, *p* < .04, *dz* = .26, but other comparisons were not significant, *p* > .16.

Second, the interaction between trial type and task, *F*(1, 2333) = 3.26, *p* = .07, suggested that participants were marginally quicker to categorise powerful individuals on males-top trials, (*M* = 1001ms, *SE* = 44.41, 95%CI[950, 1127]), as opposed to powerless ones on males-top trials (*M* = 1080ms, *SE* = 44.53, 95%CI[1000, 1177]), *t*(89) = -1.80, *p* = .07, *dz* = .19.

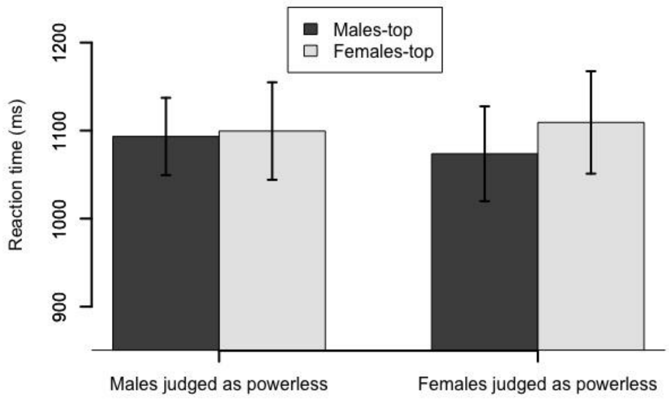
Finally, the three-way interaction between gender choice, trial type, and task was not significant, *F*(1, 2242) = .83, *p* = .36 (see Appendixes Figure 5).

Figure 5. Response latencies to (a) males and females judged as powerful, and (b) males and females judged as powerless according to the trial type (males-top/females-bottom or females-top/males-bottom). The error bars show -/+1 standard error.

(a) Powerful condition



(b) Powerless condition



Study 6: Results

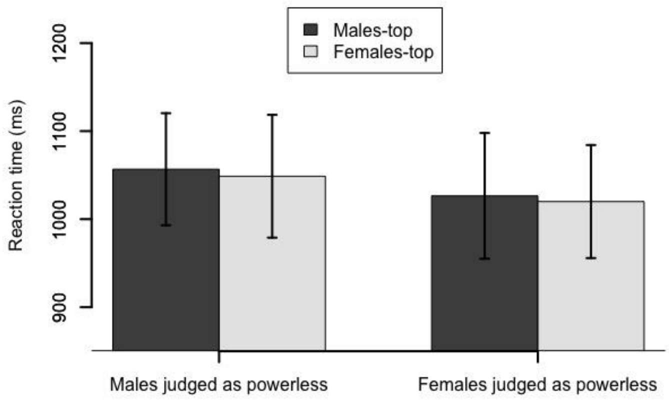
Spatial Task. Proportions. As in the previous studies, participants were more likely to indicate males as powerful (59.3%) than females as powerful (40.7%), $\chi^2(1) = 85.87$, $p < .001$, odds ratio $_{\text{choosing males as powerful versus females}}$ = 1.49 (95%CI[1.34, 1.63]).

Spatial Task. Response latencies. I found that the fixed effects of gender choice, trial type, and task were not significant, *ps* > .10. The same was true for two-way interactions, *ps* > .68, except for the interaction between gender choice and task, which was marginally significant, $F(1, 2196) = 3.24$, $p = .07$. Post-hoc comparisons indicated that participants were significantly faster at categorising females as powerless, (M = 1024ms, SE = 71.65, 95%CI[897, 1184]), than males powerless, (M = 1052ms, SE = 72.41, 95%CI[952, 1241]), $t(2194) = -2.45$, $p < .01$, $dz = .14$. Other comparisons were not significant, *ps* > .85.

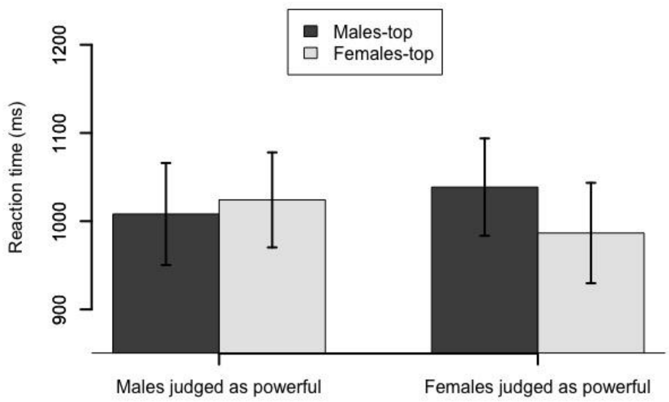
Finally, the predicted interaction among gender choice, trial type, and task was not significant, $F(1, 2176) = .20$, $p = .66$ (see Appendixes Figure 6).

Figure 6. Response latencies to (a) males and females judged as powerful, and (b) males and females judged as powerless according to the trial type (males-top/females-bottom or females-top/males-bottom). The error bars show -/+1 standard error.

(a) Powerful condition



(b) Powerless condition



Rationality/emotionality and gender IAT. Studies 5 - 6

IATs: In Study 5, I found that participants were quicker at sorting items associated with males and rationality (*M* = 716ms, *SE* = 15.58, 95%CI[685, 747) than females and rationality (*M* = 873ms, *SE* = 20.05, 95%CI[833, 913), $t(59) = -9.78$, $p < .001$, *dz* = 1.25. Similarly, in Study 6, participants were significantly faster at categorizing items associated with males and rationality (*M* = 678ms, *SE* = 10.63, 95%CI[657, 700) than females and rationality (*M* = 809ms, *SE* = 14.86, 95%CI[779, 838), $t(59) = -10.98$, $p < .001$, *dz* = 1.48.

Correlations: I calculated facilitation scores based on participants' reaction times (mean RTs to males/rationality were subtracted from mean RTs to females/rationality, i.e., the higher the score, the higher the tendency to associate males with rationality). In Study 5, I did not find significant correlations between simulation scores to powerful-males-top, $r(29) = .29$, $p < .13$, or to powerless-females-bottom, $r(28) = .15$, $p = .46$. The same was true for Study 6 (powerful-males-top and IAT score: $r(29) = -.10$, $p = .61$; powerless-females-bottom: $r(29) = .22$, $p = .26$).

Rationality/emotionality and verticality IAT. Studies 5 – 6

IATs: In Study 5, participants were significantly faster at categorising top and rationality items (*M* = 802ms, *SE* = 19.91, 95%CI[762, 842), rather than top and emotionality ones (*M* = 875ms, *SE* = 23.91, 95%CI[827, 923), $t(58) = -3.49$, $p < .001$, *dz* = .47. In Study 6, participants were marginally faster at responding to the categories of upper positions and rationality (*M* = 768ms, *SE* = 19.04, 95%CI[730, 806) than upper positions and emotionality (*M* = 807ms, *SE* = 20.12, 95%CI[766, 847]), $t(57) = -1.87$, $p = .07$, *dz* = .27.

Correlations: In Study 5, the tendency to associate top with rationality was not related to the spatial simulation of males as powerful at the top, $r(29) = -.07$, $p = .73$, and females as powerless at the bottom, $r(27) = -.20$, $p = .32$. Similarly, in Study 6, the IAT facilitation score did not correlate with participants RTs to choices of males at the top as powerful, $r(27) = .22$, $p = .28$ and females at the bottom as powerless, $r(29) = -.07$, $p = .71$.

Appendix 7

Study 8: Four sets of assigned professions and their spatial position for each gender.[41]

Females:

Set 1.

Top	Bottom
Poppy-Professor	Emily-Architect
Olivia-Doctor	Lily-Scientist
Jessica-Architect	Chloe-Dentist
Ruby-Dentist	Alice-Doctor
Molly-Scientist	Sophie-Professor
Emily-Architect	Poppy-Professor
Lily-Scientist	Olivia-Doctor
Chloe-Dentist	Jessica-Architect
Alice-Doctor	Ruby-Dentist
Sophie-Professor	Molly-Scientist

Set 2.

Top	Bottom
Jessica-Architect	Sophie-Professor
Molly-Scientist	Alice-Doctor
Ruby-Dentist	Emily-Architect
Olivia-Doctor	Chloe-Dentist
Poppy-Professor	Lily-Scientist
Sophie-Professor	Jessica-Architect

[41] The same pairs of names were presented in Study 9, however, they did not include professions.

Alice-Doctor	Molly-Scientist
Emily-Architect	Ruby-Dentist
Chloe-Dentist	Olivia-Doctor
Lily-Scientist	Poppy-Professor

Set 3.

Top	Bottom
Ruby-Dentist	Lily-Scientist
Jessica-Architect	Alice-Doctor
Molly-Scientist	Emily-Architect
Olivia-Doctor	Sophie-Professor
Poppy-Professor	Chloe-Dentist
Lily-Scientist	Ruby-Dentist
Alice-Doctor	Jessica-Architect
Emily-Architect	Molly-Scientist
Sophie-Professor	Olivia-Doctor
Chloe-Dentist	Poppy-Professor

Set 4.

Top	Bottom
Ruby-Dentist	Sophie-Professor
Olivia-Doctor	Emily-Architect
Poppy-Professor	Alice-Doctor
Molly-Scientist	Chloe-Dentist
Jessica-Architect	Lily-Scientist
Sophie-Professor	Ruby-Dentist
Emily-Architect	Olivia-Doctor
Alice-Doctor	Poppy-Professor
Chloe-Dentist	Molly-Scientist
Lily-Scientist	Jessica-Architect

Males

Set 1.

Top	Bottom

Oliver-Professor	Harry-Architect
Jack-Doctor	Lucas-Scientist
Thomas-Architect	Ethan-Dentist
Jacob-Dentist	Daniel-Doctor
James-Scientist	David-Professor
Harry-Architect	Oliver-Professor
Lucas-Scientist	Jack-Doctor
Ethan-Dentist	Thomas-Architect
Daniel-Doctor	Jacob-Dentist
David-Professor	James-Scientist

Set 2.

Top	Bottom
Thomas-Architect	David-Professor
James-Scientist	Daniel-Doctor
Jacob-Dentist	Harry-Architect
Jack-Doctor	Ethan-Dentist
Oliver-Professor	Lucas-Scientist
David-Professor	Thomas-Architect
Daniel-Doctor	James-Scientist
Harry-Architect	Jacob-Dentist
Ethan-Dentist	Jack-Doctor
Lucas-Scientist	Oliver-Professor

Set 3.

Top	Bottom
Jacob-Dentist	Lucas-Scientist
Thomas-Architect	Daniel-Doctor
James-Scientist	Harry-Architect
Jack-Doctor	David-Professor
Oliver-Professor	Ethan-Dentist
Lucas-Scientist	Jacob-Dentist
Daniel-Doctor	Thomas-Architect
Harry-Architect	James-Scientist

David-Professor	Jack-Doctor
Ethan-Dentist	Oliver-Professor

Appendix 8

Paired gendered names presented in Study 11.

Ethan	Zoe
Michael	Poppy
Lucas	Grace
Ellis	Evelyn
Matthew	Phoebe
David	Sarah
Nathan	Jessica
Arthur	Nancy
Lewis	Rose
Toby	Emma

Appendix 9

Table 4. Correlations between questionnaire items (Study 14)

		F-L	F-C	F-W	F-F	M-L	M-C	M-W	M-M
F-L	Correlation	1.00	.53	.79	.43	.39	.29	.44	.19
	p-value	.001	.001	.01	.01	.04	.01	.19	.19
F-C	Correlation	.53	1.00	.44	.21	.32	.46	.39	.22
	p-value		.00	.14	.02	.001	.01	.14	.14
F-W	Correlation	.79	.44	1.00	.31	.36	.35	.45	.05
	p-value	.00		.03	.01	.01	.001	.73	.73
F-F	Correlation	.43	.21	.31	1.00	.34	.14	.23	.30
	p-value	.14	.03		.02	.33	.10	.04	.04
M-L	Correlation	.39	.32	.36	.34	1.00	.63	.81	.48
	p-value	.02	.01	.02		.00	.00	.00	.001
M-C	Correlation	.29	.46	.35	.14	.63	1.00	.61	.47

	p-value	.00	.01	.33	.001		.001	.001	.001
M-W	Correlation	.44	.39	.45	.23	.81	.61	1.00	.28
	p-value	.01	.001	.10	.00	.00		.05	.05
M-M	Correlation	.19	.22	.05	.30	.48	.47	.28	1.00
	p-value	.14	.73	.04	.00	.00	.05		

Note: F-L: Female faces liking ratings; F-C: Female faces competence ratings; F-W: Female faces warmth ratings; F-F: Female faces femininity ratings; M-L: Male faces liking ratings; M-C: Male faces competence ratings; M-W: Male faces warmth ratings; M-M: Male faces masculinity ratings.