Role model effects on women's political engagement

Observational and experimental approaches to measurement & two studies on mediation.

Submitted by *Laszlo Horvath* to the University of Exeter
as a thesis for the degree of

Doctor of Philosophy in Politics,

on 23 March 2018.

Advisors:

Prof. Susan Banducci, Department of Politics
Dr. Samuel James Vine, College of Life and Environmental Sciences

"I certify that all material in this thesis which is not my own work has been identified
and that no material has previously been submitted and approved for the award of a
degree by this or any other University."

Signature

This page intentionally left blank.

Abstract. Contributing to a growing debate about 'symbolic' or non-policy effects of gender-balanced legislatures, my thesis sets out to tackle issues of (a) measurement, combining experimental and observational evidence of the effect of female politicians as role models on women's political engagement; and (b) mediation, considering the underlying mechanisms convincing on the individual-level of voter psychology, explaining *why* role models are powerful in engaging fellow women in the electorate.

Firstly, I triangulate results from an eye-tracking experiment investigating attentional bias to gender balance in manipulated picture stimuli of political groups; an online experiment investigating measures of psychological engagement with politics as a function of gender balance in the same picture stimuli; and British Election Study panel data investigating campaign effects on psychological engagement with politics as a function of the gender balance among candidates running in the 2010 and 2015 UK parliamentary constituencies.

My results suggest two general types of role model effects: one of 'tokenism' where women's striking minority presence impacts political attention and the probability of learning about politics, and one of 'linear' effects where a gradual increase in women's presence in political groups towards parity translates into a gradual increase in political self-efficacy and confidence about political knowledge.

Secondly, I develop and test hypotheses about mediation in terms of implicit mechanisms not requiring that citizens consider the policy output of their representatives, drawing heavily on the stereotype threat literature especially on the role of affect. Using a more classical, regression-based approach to mediation analysis, along with a novel crossover experiment or 'design-based' mediation analysis, I present preliminary evidence that, following exposure to role models, women experienced fewer self-evaluative threats as evidenced by anxiety, explaining effects on self-efficacy in politics. I present an additional study scrutinising affect, and show that the action-oriented anger may result in approach of the source of threat, reversing stereotype threat effects under 'men-only' politics.

Thirdly, I develop and test hypotheses about mediation in terms of instrumental mechanisms that do require expectations or associations about policy output. Through similar approaches to mediation analysis, I show that though women expect better policy across two domains with more female politicians on board, greater competency attributed to elites is, if anything, negatively related to self-efficacy in politics. In a full-experimental study, I find no evidence that women's greater self-efficacy is due to expectations about women-friendly policy pursued by role models.

This page intentionally left blank.

This page intentionally left blank.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

> To every little girl who dreams big: Yes, you can be anything you want—
> even president. Tonight is for you. -H [1]

Many scholars have focused on the policy consequences of women's underrepresentation in political life, suggesting that women friendly policies are more likely to be enacted when more women hold political office. More recently, research has turned to exploring 'symbolic' or non-policy effects of more gender balanced legislatures, to investigate an empowering or 'role model' effect increasing women's political engagement. This growing body of research has found that where more women hold or contest elective office, women have greater stores of political knowledge; especially young girls are more likely to engage in political discussion; and some of these studies report average effects on political interest, political efficacy, and trust, though failing to replicate in some studies.

Yet a lot of this evidence relies on observational data, making it difficult to disentangle whether contexts where role model effects are detected are simply cultures where women are more politically engaged. While relevant experimental evidence exists, it is reported in isolation and outside the context of political engagement as a dependent variable. Studies utilising exposure to varying gender balance on the group level, such as a cabinet or a national parliament, rather than individual candidate gender, continue to be relatively rare. Perhaps more importantly, the majority of studies fail to address important questions about the underlying mediating mechanisms to explain *why* political role models are powerful on the individual level. With the exception of a general hypothesis about cost-benefit calculations associated with political participation, we currently lack well defined or competing hypotheses on this level.

My thesis sets out to address these gaps by designing a series of studies taking up issues of measurement as well as mediation. In the context of my work, 'measurement' refers to a more systematic investigation of cause and effect where observational data is combined with experimental data, and to reporting a wide range of indicators of women's political engagement across all these approaches. 'Mediation' refers to testing an elaborate causal chain where I identify interim variables that explain, rather than condition,

control or moderate, why exposure to an increasing gender balance in politics is relevant for the individual in terms of a convincing psychological mechanism.

Chapter 2 is a theoretical treatment of the study of mediation, developing two sets of hypotheses. Firstly, I draw on specialist literature on stereotype threat, an experimental paradigm developed to test the impact of harmful stereotypes on academic achievement, documenting evidence that these often translate into self-evaluative threats, damaging performance. Using extensions of the original framework across various domains, such as sports and STEM careers, I arrive at the conclusion that 'men only' politics may be a high threat environment for women, evidenced predominantly by anxiety, which may turn women away from politics. By contrast, an increasing presence of role models alleviates these anxieties resulting in greater political engagement. Secondly, I revisit more instrumental theories about mediation in which citizens form positive expectations about the policy output pursued by women in office. My own instrumental hypotheses draw on literature on stereotypes about issue competencies, on literature regarding gender diversity and public perceptions of corporate governance, as well as on the representation literature on 'woman friendly' policies.

Chapter 3, Methodology is a reference chapter of technical details about data collection, procedures, picture stimuli and notes on effective sample sizes. While describing an analytical strategy broadly, specific methods and statistical models are described in the individual study sections.

Chapter 4 sets hypotheses about mediation aside to first present evidence of role model effects themselves, where the emphasis is on observation and comparison across measurement approaches. I triangulate results from an eye-tracking experiment investigating *attentional engagement* with gender balance in manipulated picture stimuli of political groups; two rounds of online experiments investigating measures of *political engagement* following gender balance in the same picture stimuli; and British Election Study panel data investigating campaign effects on political engagement following gender balance among all candidates running in the 2010 and 2015 UK parliamentary constituencies. The dependent variable in the eye-tracking study is attentional bias to improved gender balance in pictures, measured through data on the location and duration of eye fixations and saccades. In all survey studies, dependent variables are political interest, performance on a political knowledge test, confidence and self-efficacy in politics, external political efficacy, and trust in Members of Parliament.

Gender balance or women's numerical presence is the key independent variable, or treatment, in all observational and experimental studies. In my Analytical Framework, summarised in Section 3.1, I consider the disadvantages of treating numerical balance as a single continuum ranging from 0% to 100%, which may be appropriate if assuming linear role model effects, but may also hide an important heterogeneity of effects at *discrete proportions* where scholarship especially on descriptive representation theorised that women make a difference. I will therefore introduce discrete values of numerical presence to tap role model effects beyond linearity, such as the differential impact of women's 'token' or

striking minority presence, in larger groups estimated near 20%; women's critical mass presence, in line with many studies near 30%; and a parity of men and women thus 50%, an ideal target of contemporary quota policies.

Chapter 5 is composed of three studies looking at the role of affect in evidencing and alleviating stereotype threat and role model effects, respectively, on psychological engagement with politics. In Study 1, I look at self-reported state anxiety and testing-specific anxiety regarding a political knowledge test to explore if shifts in gender balance (picture treatment) translate into shifts in anxiety levels, and whether these movements explain effects on psychological engagement in politics, as reported in Chapter 4. Study 2 is a novel, crossover experimental study where participants receive a sequence of treatments: after manipulation of gender balance in picture stimuli, I introduce negative manipulation on the mediator variable as well. The hypothesis being that role models alleviate anxiety stemming from stereotype threat. To this end, I assign a randomly selected portion of the treated sample to a 're-induced' state anxiety condition to check if role model effects diminish on psychological engagement with politics. In Study 3, I scrutinise affect by looking at emotions of approach and avoidance, experimentally induced in the eye-tracking study, to explore if the action-oriented anger about gender stereotyping reverses stereotype threat effects and results in a confronting the source of threat.

Chapter 6 is composed of two studies exploring instrumental or policy-relevant linkages between women politicians and women in the electorate. In Study 1, I test if women's policy expectations improve as a function of improving gender balance across the domains of healthcare and finance policy, and if these explain increased political engagement. These domains are meant to be general enough to measure an overall judgement of the course of policy but specific at the same time in the sense that gender stereotypes prevail in how well suited female politicians are in managing them. In Study 2, I design another crossover experiment where negative manipulation is an attempt to remove women's specific expectations that female politicians pursue policy that is woman friendly, to check if role model effects on political engagement diminish.

With these studies, my thesis aims to make a number of contributions. Firstly, to make additions to the body of evidence, with often mixed results, that non-negligible portions of the gender gap in political engagement are due to women's disproportionate representation in political office. I aim to do this on different levels of abstraction in terms of micro-level evidence of attention, on self-reported measures of psychological engagement with politics. Secondly, taking numerical balance seriously and exploring effects at discrete proportions of women's presence, I aim to present findings adding to lesser-explored benefits of policy remedies such as quotas encouraging women to take positions in private and public organisations alike. Thirdly, Chapter 5 takes up the secondary goal of extending the stereotype threat approach for the first time to the domain of political engagement, an experimental paradigm that has recently been implicated in the wake of the 'replication crisis', with researchers pointing at mixed evidence of effects across domains. Finally, Chapter 6 aims to encourage an extension of research on policy

competencies from the individual candidate to the level collective political bodies such as a government cabinet or a legislature, to study how (gender) diversity impacts group-level performance and public expectations.

Chapter 2

Mediating role model effects:
A theoretical framework

> When women look at tech companies and math departments, they see few women. This activates the stereotype that women aren't good at math. The stereotype [...] makes it harder for women to enter those fields. To stay. To thrive. [2]

> The business case for increasing the number of women on corporate boards is clear. [...] It is clear that boards make better decisions where a range of voices, drawing on different life experiences, can be heard. That mix of voices must include women. [3]

> *This chapter points to key empirical evidence behind an empowering or 'role model' effect of women's presence in diverse fields, and argues for the need for more work on the underlying mechanisms convincing on the individual level. I develop two sets of hypotheses about mediating mechanisms of symbolic representative linkage: that political role models alleviate self-evaluative stereotype threats, often revealed by anxiety; and that role models may be expected to contribute to better policy whether in generalised terms through competencies or specifically in implementing 'women-friendly' policies.*

Gender gaps in political engagement persist across time and jurisdictions. With the exception of turnout at elections in the United States [4] and most Western European democracies [5],[6], women are found to be comparatively less interested in politics [7],[8],[9], less knowledgeable about politics [10],[11], less likely to be contacted directly by campaigns [12], and less likely to consider, or to be encouraged to, run for office [13],[14].

In early comparative work, Verba, Nie and Kim [15] consider the disparity in 'psychological involvement' with politics across seven countries and find that higher education affects women's political interest more than men's [pp. 258–259], but also that men's baseline level of interest is already about half to one standard deviation above women's, making it difficult for women to 'catch up to the male level of political concern' [p.262].

Looking for answers beyond education, Welch [16] employs a more exhaustive set of controls of socio-economic status on British survey data but obtains a similar residual gap in political interest between men and women.

Though not tested explicitly, both studies conclude by tapping into psychological processes unrelated to socio-economics to explain these differences: the suggestion in Verba et al. seems to be centered on women's self-efficacy, that politics would 'be [seen] outside of the proper role of women' by women themselves [15, p.263], while Welch's to be more about women's judgements of the efficacy of politics itself 'where men are not particularly successful in solving policy problems, but women are largely shut out of attempting such solutions altogether' [16, p.44].

Crucially, Welch here points to the possible impact of Margaret Thatcher's incumbency in the UK (at the time, in office for only a year) in shifting these judgements, speculating that more women in power would send the implicit message that women too are capable of influencing politics. The first empirical tests of this in the studies of Verba, Burns, and Schlozman [7] as well as of Burns, Schlozman, and Verba [17] were affirmative: in the United States, women represented by female Senators performed better on tests of political knowledge, reported to be more politically efficacious, and more politically interested (in latter but not former report).

Later labelled as the role model effect, studies have followed finding positive associations between the share of women MPs and respondents' frequency of political discussion cross-nationally [18]. Another study reports that adolescents' anticipated political involvement increased following a surge in the number of women candidates running with high media visibility, over time, through 25 years [19]. While shifting focus from psychological engagement to participation, these studies further support the hypothesis that women's improving numerical balance in politics, or 'descriptive representation', may eventually close the gender gap in political engagement, maintaining that the relationship is causal [20],[21],

In so doing, they propose the first hypotheses on mediation or underlying mechanisms giving substance to the 'blank' causal link between female representatives citizen engagement. These include shifting perceptions about the appropriateness of politics for women, and policy responsiveness. State of the art explanations, however, do only half the job to link women's numerical representation directly to citizen engagement, or conversely, their minority position to citizen disengagement. While one observation is that women candidates stir political discussion in the family [19], in turn mediating role model effects on young girls' political ambition, further links between the presence of women as candidates and parents' propensity to bring up political topics at home are left untheorised.

Looking at the idea of mediation in a wider context of numerical representation of racial and ethnic minorities, many hypothesize some aspect of psychological engagement with politics, such as political interest or political efficacy, to 'mediate' effects on mobilisation [22]. There is, however, little to no explanation provided on why exactly a

descriptive representative would increase psychological engagement in the first place, being a dependent variable in its own right. In the same study on mediation, for instance, perception of 'government responsiveness' is studied as mediator to infer on some sort of cost-benefit calculations associated with women politicians, as the authors intended, itself however may be broader than that and could be treated as lagged dependent variable on another aspect of political engagement namely external political efficacy or trust.

Other than a need for a more complete theoretical account, more work on mediating mechanisms may also be useful to understand sometimes contradictory evidence behind the role model effects of female politicians in empirical investigations in the past twenty years. While results converge on a positive impact on *political knowledge*, women being more likely to recall political information if living in areas where high visibility offices are held or contested by women [7],[23],[17],[24],[25][1], and especially on young girls' frequency of *political discussion* within networks of friends and family [19],[18],[25],[21], studies seeking to isolate effects on *political interest* and *political efficacy* offer mixed evidence at best [24],[26],[27]. While some work explains why role models *do not work* by pointing at moderating factors such as partisanship [28],[29], understanding why role models *do work* when they do may have important implications for where to look and what kinds of effects to expect through what kind of exposure.

In this chapter, I introduce elements of a theoretical framework to study mediating mechanisms that are convincing on the individual level. As a point of departure, I consider an apparent dualism of explanations in the literature briefly reviewed above: one stream of theorists engaging with the role of less tangible and *implicit* concerns about women's self-efficacy in the masculine domain of politics [17],[15]; whereas another stream highlighting more *instrumental* concerns about inadequate or irresponsive policy [16],[22] including Jane Mansbridge's influential ideas about communication barriers between male politicians and female citizens, harming substantive representation [30].

While both make sense broadly as components of a causal chain linking representatives with their constituents, in their current form they also face limitations. On the one hand, their role in the representation literature has largely been theoretical, merely motivating studies that seek to detect role model effects, themselves however untested as mediators. On the other hand, as theoretical constructs they are not specific enough. For example, significant work in social psychology points to a range of potentially relevant 'automatic' or consciously inaccessible responses to stereotypes, including examples of cognition such as activation of associations [31], examples of affect such as anxiety [32], or physiological responses such as startle eye blink [33].

The aim of the following sections is to highlight specialist literature that will be used to develop ideas about mediation in response to these types of mechanisms, namely those more implicit in the sense that they do not require that citizens judge representatives consciously regarding their policy output, and those more instrumental that do require that expectations are formed about policy output[2].

The first major section introduces literature on stereotype threat mainly to arrive at af-

fect, and predominantly anxiety, as a potential mediator of political (dis)engagement under political (mis)representation. It develops a set of hypotheses in which no or minimal levels of representation are high threat and high anxiety contexts harming self-efficacy assessments (see especially Section 2.1.1) in politics and result in political disengagement. By contrast, when equipped with an increasing presence of role models, politics does not induce self-evaluative threats, and the result is greater political engagement.

The second major section focuses on literature about perceptions as well as stereotypes about women's competence in policy, to suggest that citizens may expect different policy outcomes from groups where women contribute in at least two important ways: first it may be assessed holistically 'better', showing diversity of expertise and viewpoints, and second, it may be expected to deliver policy that is 'woman friendly'.

There are two important limitations to highlight. Firstly, these mechanisms are meant to give an account on why role models are powerful on a personal level. When developing hypotheses about mediation, I do not move on to consider broader and more elaborate mechanisms that involve external actors and institutions such as the media that may be useful for a full account on how role models get their message across, that are beyond the current aims of exploring convincing individual-level mechanisms. Secondly, this chapter does not set out a full account on role model effects are mediated. Within implicit mechanisms, there should be a large number of alternative or parallel ideas that may explain a great deal about voter psychology under different representation setups, and within instrumental considerations, there should be further policy areas worth exploring to understand what exactly is expected from the tenure of role models. The current ones are demonstrative examples of the kinds of literature that future studies could draw on, aiming to explore the usefulness of the implicit-instrumental distinction in the first place.

2.1 Stereotype threat and affect

Steele and Aronson define [32] stereotype threat essentially as 'inferiority anxiety' about confirming a negative stereotype, which may arise in specific situations where "negative stereotypes provide a possible framework for interpreting [one's own] behaviour" [34, p.415]. Drawing on early research on the underachievement of Black people in academia, the authors show that the achievement gap can be reinforced or closed by altering participants' threat-perception with very explicit cues. In this first investigation of stereotype threat effects, experimental manipulation *induces threat* by informing a sample of participants about Blacks' underachievement on tests such as the upcoming one, relative to Whites. Other participants receive treatment *pre-empting threat* by assuring them about no race differences on tests such as the upcoming one.

Studies following this experimental paradigm show that those reminded of harmful stereotypes—most often race [35],[36] but extended to gender [37], race *and* gender [38] and age [39] related stereotypes as well—perform worse than those who were not, suggesting that an achievement gap in education and careers may be at least partially due to a series of such responses to stereotypes encountered and internalised through the life

course. While robust across race and gender, studies typically find moderate effects when working with stereotypes regarding women's quantitative skills, however, meta-analysis points out that women predictably react to even subtle cues whereas Blacks tend to need very explicit reminders of stereotypes to 'open up' the achievement gap [40].

In the present theoretical framework, two extensions of the stereotype threat literature are used to strengthen the role model hypothesis about women's symbolic role in political office, and a third in particular to develop a hypothesis about mediation that is better suited for empirical testing. Next, I will briefly review some of the work done on the effects of stereotyping *beyond test performance*, arriving at the broader concept of self-efficacy, suggesting that this framework may be used to study women's psychological engagement with politics. Then, I will review experimental work where gendered stereotype threats were successfully alleviated via exposure to visible *role models* in a number of fields such as STEM careers, business, and sport, informing hypotheses about the potential effect of numerical balance in political elites. Finally, I develop ideas about mediation using the role of *affect*, most prominently anxiety, making stereotype threat-related mechanisms detectable.

2.1.1 Effects beyond testing: self-efficacy

Studies synthesising the first results of stereotype threat experiments conclude that a series of negative experiences under racist stereotype threats ultimately impacts one's intellectual identity, which may result in academic disidentification altogether (e.g. "Studying is not for me") [41],[42].

In terms of academia and jobs, studies tend to hypothesize partial identification or domain selectivity when looking at gender stereotypes specifically. A novel experimental design finds that its women participants subjected to sexist television commercials were consequently less likely to express a preference for careers relying heavily on a quantitative skill set and preferred careers relying on 'verbal domains' such as creative writing and communications [43]. Women's disproportionate presence in STEM careers remains the focus of ongoing research [44],[45],[46], drawing heavily on stereotype threat scholarship.

Behind these instances of domain avoidance, common theory is that stereotypes harm individuals' *self-efficacy* judgements ("Am I able to succeed?"), resulting in a move away from the source of threat, though the specific mechanism is still debated [47]. A definition of self-efficacy in politics is less straightforward, but it has its parallel: election studies often look at 'internal political efficacy', a construct understood partly as *confidence about one's political awareness* on one hand, and as *judgements about the benefits of voting* and other participatory efforts, on the other [48],[49]. Applying theory from the extended stereotype threat paradigm on these constructs, one may hypothesize that masculine politics induces self-evaluative threats in contexts where women need to assess their own political skills, a typical solution in turn being political disengagement.

There is no direct application of the stereotype framework on this or any other as-

pects of psychological engagement with politics, such as political interest or trust, but two studies come close by looking at simple performance on political knowledge tests. Relatively recent study demonstrates interviewer gender effects on how participants of a telephone survey performed on a political knowledge test [50]. The finding, that women scored higher when interviewed by women, is consistent with the possibility of stereotype threat effects on confidence, if a correct answer is driven by a greater confidence in one's own knowledge instead of opting for refusal or selecting 'don't know'—important work on political knowledge points at a gender gap between men's and women's propensity to guess on these tests [51]. Another, unpublished study looks at similar measures and finds that the gender gap in test performance is cut by half if participants previously rated a mix of male and female politicians, rather than only men, on performance thermometers [52]. Another relevant study looks at political ambition, whether women anticipate to run for office, finding weak evidence of a stereotype threat effect [53].

2.1.2 Role models alleviate stereotype threat

Going beyond the causes of stereotype threat, studies quickly follow up showing ways that stereotype threats may be successfully alleviated. One such proposition is about exposure to a visible and competent role model [54] showing that women's test performance is 'safeguarded' by a female test administrator, with important effects on confidence about math ability, and self-esteem. In another study, participants read about gender-manipulated fictitious biographies, finding that exposure to the stories of women inventors, lawyers, architects, and doctors, without an explicit mention of their quantitative skills, increased female participants' math scores [55].

Further studies consider variation in gender balance by using visual stimuli. Further investigating STEM subject selectivity, one study reports that women viewing a (staged) academic conference video with a 1:3 ratio of women to men were less likely to express interest to participate at the conference than those viewing a parity of men and women [45]. Within STEM, recent review highlights inconsistencies across studies in the correlation they report between the proportion of female science faculty and the proportion of women majoring in science, suggesting that it may a case of non-linear relationship where a critical mass of women teachers may matter but beyond such a threshold, marginal effects diminish [46, p.13].

While political engagement has not yet been implicated in these studies as a dependent variable, exposure to political role models has been considered to alleviate stereotype threat in other domains. Most notable is the 'Obama effect'[3] coupled with some evidence that Barack Obama's visibility and novelty in the highest executive office improved GRE-style test scores, followed on a small rolling cross-section of Black participants [56]. This finding, however, is not mirrored in a parallel laboratory experiment, if not casting doubt on Obama's power as a role model altogether but qualifying it as 'less pervasive and more dependent on test-taker characteristics' [57, p.960].

First evidence of a more complex 'Hillary Clinton effect' comes from a study [58] that

starts with a free name recall task asking a pilot sample of students to list famous women who deserved success, understood broadly to include celebrities, and famous women who did not deserve success, and finding that Hillary Clinton is the only name listed top 3 on both lists. Thus the researchers proceed to design the main study in which participants are asked to rate Hillary Clinton on 'deservingness', and are then presented with her biography months later as treatment before taking a GRE test, demonstrating that role model effects depend on initial ratings.

Another, novel study used picture stimuli while asking participants to deliver a speech in which they took position against an increase in tuition fees [59]. Speeches were delivered in a Virtual Reality environment depicting an audience seated in a room either without a framed picture on the wall, or one featuring Bill Clinton, Hillary Clinton, or Angela Merkel. Looking at both objective measures of speech quality in terms of length and sophistication, and a subjective measure of performance, there were notable gender gaps both in the control and the Bill Clinton groups, but none under either role models. Women in the Merkel condition did, however, markedly outperform other women in the Hillary Clinton condition.

Entirely outside of the stereotype threat framework, one study re-examines individual women politicians: Hillary Clinton, Nancy Pelosi, and Sarah Palin, by looking at the effect of their media presence longitudinally on young women's 'anticipated political involvement' including intention to campaign or donate, and finds only subgroup effects among Democratic identifiers [60]. Looking at the three investigations together, they may serve as reinforcement of the Obama effect in that individual politicians enter the public arena with a number of relevant attributes including personal stories, partisanship, and contextual characteristics, conditioning or moderating role model effects.

More optimistic results come from studies using non-American political role models. A very high external validity natural experiment in India reports that one of its state's gender quota policies, whereby a randomised third of all village councillor positions were contestable by women only, increased young girls' career ambition and efficacy in numerous fields including leadership roles [61]. Another paper uses the same policy in West Bengal and Rajasthan to show important effects on women's political participation, in terms of contacting the councils as well as showing that policy preferences shifted in complex ways, but generally away from supporting men's work to welfare provision in rural areas [62]. However, randomised quota policies failed to achieve these ends in other reports. In Leshoto in Southern Africa, quotas *turned women away* from participation, which the authors explain by referring to public opposition to quotas [63]. A study looking at Latin American quota effects finds small increase in women's political interest but not on other indicators of political engagement, although fails to control for individual-level covariates such as age or education [64].

Parallel studies, especially in education and business psychology, report role model effects derived from female instructors or supervisors without relying on the stereotype threat literature explicitly. One such stream of research uses Social Learning as theo-

retical background, putting emphasis on imitation of behaviours both of mentors and of role models [65],[66] though some appeal to common sense notions of role models challenging stereotypes [67]. Perhaps worth noting from this field, an observational study on students' module taking behaviour that finds *negative* role model effects in some subjects, including political science and natural science, confirmed positive role model effects on mathematics and statistics subjects [68].

It is thus possible to detect, or to fail to detect, role model effects with or without using the original stereotype threat framework. While many studies referred to role models alleviating threat, others used only a general notion of stereotypes, or used different theoretical background not mentioning psychological responses to stereotypes. *Negative* role model effects were explained by moderators such as leader evaluation ('deservingness'), social distance, or partisanship. Some theoretical treatment outside of the scope of this review implicated harmful or self-deflating social comparisons as possible mechanisms behind negative effects [69],[70]. Since my argument is about mediation using stereotype threat, I will now seek ways to obtain a better understanding of such self-evaluative threats beyond merely hypothesizing their presence, with the ultimate aim to consider measurement to infer stereotype threat, in the next chapter. In order to achieve this, the next paragraphs turn to describing previous research that used affect to derive stereotype threat and role model effects.

2.1.3 Affective components of stereotype threat

While stereotype threat is used as a possible framework to explain women's political (dis)engagement under different representation setups, ranging from men only politics to women's token, critical mass, and parity presence, stereotype threat theory itself does not readily translate into a tangible, measurable mediating mechanism that may be evidenced by survey or physiology data, or other behaviour. The basic paradigm merely infers stereotype threat, by observing differences in performance, subject to the saliency of stereotypes. A number of studies, however, make advances in pinpointing elements of stereotype threat that are used to develop testable hypotheses regarding mediation.

Anxiety has been central to theorising about stereotype threat. Self-reported state anxiety (feeling worried, nervous, indecisive at the moment) is pointed at as a key 'mediator' of stereotype threat effects themselves in an early study on women's math performance [37], the central idea being that threat in the environment translates into greater anxiety, which decreases test performance. Follow-up studies replicate this mechanism under gendered, racial, and age-based stereotype threats [71],[39] but a number of parallel attempts fail to detect self-assessed anxiety mediation in 'old' [72] and 'new' [73] contexts.

Further experimental investigation looks beyond self-reported anxiety. A novel study using stereotypes about sexual orientation and suitability for childcare finds that while its survey measures capture no anxiety effects, external judges monitoring nonverbal behaviour do find anxiety differences between those in the threat and those in the control conditions, in turn explaining effects on childcare performance [74]. A cognitive measure

investigated under exposure to gender stereotypes is vigilance, a state of alertness, with a study finding that those who viewed videos of gender unbalanced panel discussions with only a token presence of female panelists recalled a greater number of details about their surroundings and details from the videos as well [45], a phenomenon with explicit links to anxiety, e.g. [75], and with 'political concern' being also linked to anxiety in political behaviour [76].

A study on racial stereotypes finds short term (up to 20 minutes) arterial blood pressure differences between stereotyped and nonstereotyped groups, linking these movements to mathematics test performance [77]. Though still indirectly, these results are suggestive of persistent anxiety effects: short term spikes in blood pressure have already been pointed at as likely markers of anxiety [78]. Similar dynamics are later reproduced on a cognitive ability test but using heart rate variability [79], showing activation in stereotype threat conditions. Another study finds women subjected to stereotypes have elevated cortisol levels [80], treated as biomarker of stress, which in turn may be indexed by anxiety, as the authors note. It is thus possible that though *in some contexts, anxiety is salient and accessible* for participants to report in a survey, *in others it may not be*.

An important difference between the above measures and self-report, other than accessibility, is *timing*. The above studies detect anxiety effects during task completion, while survey items are typically asked in isolation. While a seemingly methodological consideration, timing may have substantive implications on the *kind of anxiety* researchers are capturing. A relevant stream of literature distinguishes diffuse state anxiety approaches to measurement from more situation-specific anxieties [81], the agenda including the development of inventories developed specifically for testing anxiety [82] that proves better at predicting performance on standardised tests, or other performance anxieties such as 'sport anxiety' [83] proving to be better at predicting sport performance.

It is thus possible that mixed evidence of anxiety under stereotype threat is partially due to some studies trying to capture more *diffuse state anxiety* while others capture *situation specific anxiety*. A very recent study makes this differentiation explicit within the stereotype threat framework as well [84]. In the framework proposed for my analysis, the two are treated as different approaches to arrive at a singular underlying concept, namely anxiety, and ask first if they describe individual experiences under varying numerical representation setups, and second if dynamics in anxiety experiences explain political engagement.

[H1.1] *As the proportion of role models in political office increases, women experience fewer self-evaluative threats in the political domain, evidenced by decreasing anxiety.*

[H1.2] *In the absence of self-evaluative threats in the political domain, decreasing anxiety improves psychological engagement with politics.*

Anxiety's prominent place in the literature may be due to its extensively researched link with testing. Outside of the anxiety canon, however, one study seeks to explore po-

tential diversity of negative emotions relevant for stereotyping [85] based on first-person accounts (a 'thought listing' exercise) of 60 female psychology students taking a difficult mathematics test, finding that (a) anxiety and uncertainty about self-efficacy in math prevail and that (b) at least hatred and extreme aversion of math may be ruled out as typical emotion under stereotyping.

Those extremities are, however, observed with a time lag in recent study on 'stereotype spillover' [86]. Here the authors investigate what happens *after* putting someone through gendered stereotype threat, and find that while women provided with adaptive coping strategies showed no discernible character in a subsequent cooperative game, those short on adaptive coping strategies showed heightened aggression as typical gaming strategy.

Similarly, further research implicates anger when it comes to stereotypes but directed at the outgroup: the source of the stereotype (or the source of threat). Importantly, political psychology literature shows that anger drives protest intention as a function of identification with the ingroup, which in turn may be increased or decreased by altering threat perception [87]. An influential study on party identification shows that a limited number of identifiers continues to rely on programmatic considerations to support campaign involvement, but there is a considerable element of 'expressive partisanship' of which action-oriented emotions, such as anger, are the best predictors [88].

Thus while the majority of studies hypothesise anxiety and disengagement under stereotype threat, alleviated by relying on role models, there may be greater individual heterogeneity in emotional reactions where anger is a reaction to stereotype threat which directly motivates *more* political engagement. In this case, the emphasis is not so much on the traditional dichotomy of anxiety and anger as discrete emotions, but on opposing ends of a more continuous affective dimension that differentiates how people cope with a negatively-valenced situation [89], namely an action-oriented option and an avoidance-oriented option. A similar idea is pursued by research in neuroscience looking at approach and withdrawal processes, often explicit about affect [90].

While not forming hypotheses about why some people adopt one coping strategy over another, I do consider if such coping strategies exist and whether they make a difference in terms of political engagement: do those showing anger over stereotypes in the political domain engage with (confront) politics? Do either action-oriented emotions, such as anger, or emotions of avoidance, such as helplessness or sadness, have causal primacy in this mechanism?

[H2] *Emotions of sadness and helplessness motivate a move away from the source of threat and disengagement from politics.*

[H3] *Action-oriented emotion such as anger motivates approach of the source of threat and increase political engagement.*

20

This concludes my consideration of stereotype threat and affective mechanisms as mediators of stereotype threat, and female politicians' role model effects on women citizens' political engagement. I will restate my hypotheses more formally, with a clear map of measurement strategy at the end of Chapter 3. My studies of mediation investigating support for these hypotheses are all conducted in Chapter 5. In the remainder of this chapter, I move on to develop an instrumental link between women as descriptive representatives and women as constituents.

2.2   Policy competence

Despite its decreasing prominence in recent electoral research, evidence is robust, if context-dependent, regarding instrumental or policy-based linkages between parties and supporters. In mature democracies, still large segments of voters support especially well-established parties based on their policy platforms [91], the news media successfully facilitates issue voting for those attentive [92], and relatively new experimental evidence at least partially restores faith in the stability of citizens' issue positions [93].

In these studies, instrumental links are hypothesised when looking at party support as a dependent variable, or voting more specifically. Empirical models of voter turnout on the individual level have long been considering ways perceived benefits may outweigh the costs associated with voting, the instrumental option being policy return [94]. At least in the political science mainstream, however, studies have not engaged with the possibility of an instrumental or policy-based aspect of psychological engagement with politics, such as political interest, trust, or political efficacy—with the notable exception of scholarship on 'issue publics' [95] who seek political information selectively based on unique policy concern [96].

There is more about instrumentalism in the symbolic representation literature. Seminal study on Black empowerment extends the instrumental argument from participation to trust in government, political efficacy, and political awareness, all derived from cues Black citizens take from Black policy makers [22]. These expectations are more detailed in theoretical work on women's representation. Though both normative and empirical work finds limited evidence that 'women act for women' in American government, the instrumental hypothesis is that women policy makers open up communicative channels where citizens' policy preferences are finally articulated replacing 'mistrust', and that women policy makers may contribute to the crystallisation of policy preferences (of women) that were thus far not politicised by political elites (men) [30].

Direct support for this hypothesis comes from recent work on newly elected women village councillors in India receiving and tackling new policy complaints supporting women's life and work in rural locations rather than men's work and commute to urban areas [62]. Against this participatory aspect, there is little investigation into effects on psychological engagement with politics, except an American survey study finding that role models may not have impacted women's perceptions about government responsiveness [18].

In the next sections, I elaborate on two ways instrumentalism, that is, a consideration of policy output, may impact women's political engagement while remaining applicable to various aspects of psychological engagement with politics as well, such as political interest, trust and external efficacy, and self-efficacy. While termed instrumental for being policy-relevant considerations, they do not necessarily contrast with 'implicit' mechanisms listed in the previous sections in terms of how readily and consciously accessible they are to voters. First, I develop a hypothesis that politics with role models may signal a mixture of competencies that lead to *better policy output*. Competence judgements about political figures are often formed spontaneously [97] but citizens, especially historically excluded segments, may be explicitly concerned about inclusivity, which goes beyond spontaneous judgements. Within this broad expectation about policy output, I also consider if expectations differ across issue areas where stereotypes about competency either advantage or disadvantage women. Second, I look at the specific idea from the representation literature that role models are expected to pursue *policy that is women-friendly*, to be able to form a hypothesis that an expected women-friendly output motivates greater political engagement.

2.2.1 Policy output, competencies, and stereotypes

There is a wealth of specialist literature on whether and when stereotypes are triggered, and 'applied' when voting for, a female candidate. Voters associate feminine *trait stereotypes* such as compassion or warmth with female candidates to limited degree only [98] and with limited electoral consequences [99],[100] that often prove difficult to disentangle from cross-cutting effects of partisan stereotypes [101],[102]. More consistent are findings on feminine *issue stereotypes*, where women are seen more competent managing 'compassion' issues such as health care, education, or social security, as opposed to finance, military, foreign policy [103],[104],[105],[106], suggesting that the electorate may use gender cues when evaluating performance in office as well.

Findings about issue stereotypes are suggestive that (a) when moving away from the individual candidate to politics as a group of actors, such as the governing Cabinet or the national legislature, gender mix may signal a mixture of policy competencies and priorities; and that (b) looking at specialised office such as healthcare, citizens may use gender to infer on the quality of the policy output. The next paragraphs elaborate on these ideas and specify if policy expectations could be unique to women and men in the electorate.

While studies often assess attitudes about women in politics in a 'battle of sexes' manner on a single domain of competency or a single office (e.g. 'Are men or women better suited as political leaders?', 'Would a female president anger you?'), few studies look at group-level assessments as a function of gender balance. Using 2006 American National Election Studies data, paper reports that the majority of both men and women would prefer a gender balance of around 50-50% in elective office[4] [107].

There are more measurement approaches to group- or organisation-level diversity in

recent business literature. Relevant studies on corporate board diversity find that external stakeholder perception of 'instrumental legitimacy' in terms of skills and experience increases with improved gender balance [108], that in addition to hard measures of financial performance [109], public reputation increases with gender balance [110]. Interestingly, one study breaks with the linearity assumption and finds that, across 122 stock-market companies in Vietnam, the proportion of female board members beyond 20% has decreasing marginal returns in terms of financial performance [111].

These findings on the organisational level also resonate with important philosophical position on political system legitimacy, namely its feminist critique. Iris Marion Young emphasises *inclusion* in meaningful political deliberation over gender- or race-blind democratic procedures as source of legitimacy [112]. More explicit about public perceptions about legitimacy, an experiment finds that increased descriptive representation increased African Americans' broad support of federal Judiciary in the US but not Whites' [113].

The empirical findings thus underscore that gender diversity in organisations and institutions contribute to a favourable public assessment broadly, as a holistic judgement of performance or output. Using gender as an information shortcut about policy competency, similarly broad judgements about policy output pursued by governments are now expected to improve with an increasing proportion of female politicians in them.

[H4.1] *As the proportion of women politicians in political office increases, citizens (women) expect better policy output.*

When looking at men and women's broad expectations about policy output, I do not explicitly consider gender differences. Looking back at the literature on women's (descriptive) representation, however, women's historic exclusion from political office has led many to speculate that women, more than men, are sensitive to gender balance in politics. One observation is that in the 'political status quo, [...] men are not [seen as] particularly successful in solving policy problems', which is linked causally to women's political alienation and apathy [16, p.44]. Being the dominant group, men's policy expectations may be entirely insensitive to gender balance, not having used gender as a competency cue before; or that men are sensitive to gender balance but evaluate policy more poorly, afraid of losing a dominant position.

Turning to expectations about policy in specialised office, gender-issue stereotypes may play an important role in public evaluations. It is possible that public evaluation entirely follows stereotypes, and values a greater presence of women in policy areas where women are seen competent such as health care and education. Since men and women often use similar issue stereotypes when looking at male and female politicians [114], it is likely that the hypothesis applies to both voters of both genders.

[H4.1.1] *In policy domains where women are seen as stereotypically competent, role models are expected to contribute to better policy output.*

The issue stereotype literature offers little to suggest that women in counter-stereotypic policy areas would be seen to contribute equally. Interestingly, however, an additional finding of the 2006 ANES survey study, mentioned above, is that those preferring women's presence in politics at 50% or above do give counter-stereotypic answers on some issue competency measures, namely they would prefer women managing crime [107].

Therefore, I am keeping the possibility open that role models may be expected to do well in counter-stereotypic domains as well, given earlier arguments about women's sensitivity to historic exclusion [16],[30] which may be most striking in policy areas where stereotypes do not advantage women. Thus while exploratory, my working hypothesis is that,

[H4.1.2] *In policy domains where women are not seen as stereotypically competent, role models are expected to contribute to better policy output.*

Closing the causal chain of mediation for the overarching hypothesis about policy expectations, as well as the sub-hypotheses about office type in terms of issue stereotypes, the final expectation is that better policy output links back to political engagement instrumentally.

[H4.2] *Better policy output motivates greater political engagement.*

2.2.2 Delivering a 'woman friendly' policy

Moving beyond impressions and stereotypes based on gender, I now turn to more substantive expectations about women's tenure in office drawing more directly on the representation scholarship, namely, on work engaging with the idea that women represent women's interest and issues by implementing 'women friendly' policies. While my hypothesis is primarily about the demand side namely public expectations about substantive representation, I will first briefly review approaches to what might constitute a woman friendly policy.

A catalogue of women's issues has proved to be difficult to compile. Many attempts take the gendered distribution of labour as point of departure, which in Western European countries revolved around demands in the private sphere such as fairer distribution of domestic work and caretaking responsibilities, and improving health of mothers [115]. Analysing men and women's preferences based on the content of their claims ('complaints') submitted to village councils, a study in India finds that while women rather than men raised concerns about welfare programmes, drinking water, and child care, concern about public works issues such a road improvement is rather a function of who commutes or works on those roads and thus in one state it was predominantly a 'women's issue' but in another it was a 'men's issue' [62].

Feminist approaches to women's issues emphasise equality demands and thus equal pay, reproductive rights, an end of violence against women and human trafficking [116]. Studies may frame this more broadly in terms of improving 'status' of women across a

variety of domains such as 'social, economic, and political' [117]. Contemporary political agendas such as the UK Women's Equality Party's broaden the meaning of who the constituents might be by emphasising interests of 'BAME women, disabled women, migrant women, working class women, and LGBT+ women' [118, p.9]. Another extension of the concept comes from studies recognising that a number of other actors not from traditionally left-wing politics take up women's interests. A comparative case study shows that a number of right wing actors including the extremes form arguments explicitly based on women's rights and freedoms such as in the *burqa* debate in Belgium [119].

While these insights show that women's issues depend heavily on context and politicisation, citizens' expectations, if not carefully delineated, about the representation of women's interest has been linked to women's presence in politics. Study using British Election Study data finds that while both women and men in older birth cohorts think that women are better suited to represent women's interest, the gender gap opens up in younger birth cohorts where women are more likely to think they need women MPs to be represented [120]. Another UK study reports, first, that women's highest priority in representation is that MP's represent their 'viewpoints', and second, that 59% more women think that there should be more women Members of Parliament than those who think there should be fewer, which is at least theoretically consistent with the idea that women expect women MPs to represent their viewpoints better [121].

As a specific instrumental linkage especially between women in the electorate and women policy makers might be citizens' specific expectations women-friendly output. Keeping the hypothesis broad as to accommodate a wide range of interests and priorities,

[H5.1] *As the proportion of role models in political office increases, women expect women-friendly policy output.*

while the link back to various aspects to political engagement including psychological engagement with politics is

[H5.2] *Women-friendly output motivates greater political engagement.*

2.3 Conclusions

A fundamental dynamic in political representation is that constituents retrieve some symbolic benefits from their representatives beyond a substantive representation of interests [122], which is often described in terms of community empowerment including increasing participation, or some predominantly psychological gains associated with it on the individual level. For ethnic minorities, historically oppressed or excluded groups, relatively recent democratic theory has started to outline how descriptive representation (that is, appointing representatives based on likeness, or group membership), if defined carefully, could serve the purposes of such empowerment [123, pp. 34–36].

This chapter started by considering 'symbolic' or non-policy benefits of descriptive representation in terms of political engagement, and found that a greater presence of

women as role models in political offices may positively impact not only participation but women's political interest, political knowledge, confidence and self-efficacy, external efficacy, and trust. While evidence is mixed behind each of these indicators especially due to diverging methodologies and operationalisation of descriptive representation, I found the idea of mediation to be the most striking conceptual gap: the absence of systematic investigation into why role models work, what might be a convincing individual-level psychological mechanism linking descriptive representatives to their constituents.

Drawing on a diverse set of literature that does not directly study political behaviour, I developed ideas about mediation around two types of individual-level responses to women's presence in politics, namely those more implicit in the sense that they do not require that citizens judge representatives consciously on some dimension of competency or policy output, and those more instrumental that do require that expectations are formed about policy output.

Firstly, I introduced literature on stereotype threat mainly to arrive at affect, and predominantly anxiety, as potential mediator of political (dis)engagement under political (mis)representation. It developed a set of hypotheses in which no or minimal levels of representation are high threat and high anxiety contexts harming self-efficacy assessments in politics and result in political disengagement. By contrast, equipped with an increasing presence of role models, politics does not induce self-evaluative threats, and the result is greater political engagement.

Secondly, I focused on literature about perceptions as well as stereotypes about women's competence in policy, to suggest that citizens may expect different policy outcomes from groups where women contribute in at least two important ways: first it may be assessed holistically 'better', showing diversity of expertise and viewpoints, and second, it may be expected to deliver policy that is 'woman friendly'.

These two sets of hypotheses are not meant to exhaust individual-level responses to descriptive representation, nor are they meant to be exhaustive accounts of mediating mechanisms that may include broader societal mechanisms such as the role of mass media in picking up on women's messages, or characteristics of female politicians including personality traits and experience. Rather, my aim is to demonstrate on a sample of hypotheses how specialist literature in other domains might be helpful in developing testable hypotheses thus far overlooked in the representation literature, often making only implicit statements, if any, about mediation, hypothesising rather than explaining why symbolic benefits of descriptive representation emerge.

My work proceeds as follows. After discussing an empirical strategy in the next chapter, my first empirical chapter 'Measuring the role model effect' on page 43 seeks to detect evidence that women's visibility impacts a range of psychological engagement indicators across observational and experimental approaches in a 'black box' framework that does not ask test mediation explicitly. My second empirical chapter, 'Understanding the role model effect: Stereotype threat and affect as mediating mechanisms' is composed of three studies testing the implicit link introduced above. My third empirical chapter, 'Un-

derstanding the role model effect: Expectations about policy competency as mediating mechanism' on page 124, is composed of two studies on instrumental linkage between gender balance in political groups and political engagement.

While each chapter is intended to be self-contained by drawing conclusions about the specific hypotheses tested in them, my last chapter talks directly to the dichotomy of implicit mechanisms and instrumentalism in the context of descriptive representation, as sketched briefly at the beginning of this chapter.

Notes

[1] In Reingold and Harrell's only co-partisan Role model effects were detected.

[2] My categorisation partially intersects with further dualism of categories such as expressive versus instrumental motivations behind political participation, or automatic versus controlled processes in social psychology. While it is possible to situate my proposals along these dimensions as well, it is not necessary: I can use these categories heuristically to respond to early proposals in the representation literature. Nor is it very straightforward: both these dualisms were coined having different dependent variables in mind. For instance, is is not useful to consider a linkage between women as politicians and women in the electorate as 'expressive': as noted later and especially in Chapter 6, explicit gender identification is not needed to account for role model effects. Likewise, conscious reflection on affective linkage such as anxiety is possible, as a number of studies in the stereotype threat literature used self-report or taught listing exercises to get at the role of affect (reviewed in detail later).

[3] The term may be simultaneously used to describe the president's role in decreasing an activation of racial stereotypes about Black's achievements *in out-group members* and combating prejudice. In this sense, experimental evidence is more conclusive of the Obama effect [124].

[4] Although open measures such as this are usual suspects for social desirability effects and misreporting, here the average response was 60% men in ideal government, with a left-skewed distribution, indicating minimal social desirability effects. Above in text, I reported the rounded-modal value.

2.4 List of hypotheses

The role model hypothesis

[H RM] *Role models in political office increase women's political engagement.*

    [H LRM] `Linear role model effects.` *As the proportion of role models in political office increases, women's political engagement increases as well.*

    [H DRM] `Discrete role model effects.` *Discrete proportions of role models such as a token or a critical mass of women in politics increase women's political engagment as citizens.*

Implicit linkage

[H1] *Role models in political office increase women's political engagement because they alleviate anxiety stemming from stereotypes about women's competency in politics.*

    [H1.1] *As the proportion of role models in political office increases, women experience fewer self-evaluative threats in the political domain, evidenced by decreasing anxiety.*

    [H1.2] *In the absence of self-evaluative threats in the political domain, decreasing anxiety improves psychological engagement with politics.*

Additional hypotheses on affect

[H2] *Emotions of sadness and helplessness motivate a move away from the source of threat and disengagement from politics.*

[H3] *Action-oriented emotion such as anger motivates approach of the source of threat and increase political engagement.*

Instrumental linkage

[H4] *Role models in political office increase women's political engagement because they are expected to contribute to better policy output.*

    [H4.1] *As the proportion of role models in political office increases, women expect better policy output.*

        [H4.1.1] *In political office where women are seen as stereotypically competent, role models are expected to contribute to better policy output.*

        [H4.1.2] *In political office where women are not seen as stereotypically competent, role models are expected to contribute to better policy output.*

    [H4.2] *Better policy output motivates greater political engagement.*

[H5] *Role models in political office increase women's political engagement because they are expected to contribute to women-friendly policy output*

    [H5.1] *As the proportion of role models in political office increases, women expect women-friendly policy output.*

    [H5.2] *Women-friendly output motivates greater political engagement.*

Chapter 3

Methodology

My thesis is motivated by questions about women's political engagement: In what ways does the presence of female politicians matter for addressing gender gaps in political engagement, especially psychological engagement with politics? What are the key mechanisms explaining why 'role models' are powerful on the individual level? The previous chapters theorised the role model hypothesis linking the *numerical presence of women* in visible political office causally to *political engagement* broadly, and developed new (or better specified) hypotheses about the underlying mechanisms or *mediation*.

There are a number of measurement options in terms of dependent, independent or treatment, and mediating variables. Of the following three empirical chapters, the first alone is an attempt to combine several approaches to document evidence of treatment and effect. The second and third seek ways to tease out mediation. The current chapter serves as an overview of the analytical framework and as a reference chapter of all data sources and procedures.

3.1   Analytical framework

In Chapter 2, before turning to key questions about mediation, I briefly noted limitations of the existing empirical project behind the role model effect. First, evidence is detected on one or two dependent variables in isolation, and not reliably on a larger set of relevant measures thus making reliability difficult to assess. For example, studies seeking to find effects on political discussion, or political quiz performance, report no further dependent measures of psychological engagement with politics. Second, evidence is mixed and often contradictory, without much speculation as to why role model effects implicate a particular measure of political engagement over another. Cross-country replication of results often fail as well. Third, within-subject designs such as panel studies or field experiments to improve causal inference continue to be rare. Experimental paradigms of social psychology using exposure to role models, where the focus is often on abstract measures of performance, rarely measure effects on political engagement, and offer results that are only indirectly relevant to the study of political representation. Fourth, a predominantly American scholarship on candidate gender effects tends to study single member districts

contested by few candidates only, rarely addresses questions about gender diversity in larger groups which may show more variation in numerical proportions.

My first empirical chapter, Chapter 4, aims to address some of these issues by combining observational and experimental evidence in a systematic way on key indicators of women's political engagement as well as more novel ways of measurement. Firstly, in a lab experiment, I collect micro-level evidence of political engagement, namely attentional engagement, with stimuli depicting political groups and estimate attentional bias to those that have an improved gender balance. Secondly, I designed an online experiment, and fielded twice to collect my own replication data, to estimate psychological engagement with politics evidenced by more typical survey questions, as a function of exposure to similar picture treatment in which the proportion of women varies. Thirdly, I identified the same numerical proportions among all candidates running in the 2010 and 2015 UK parliamentary constituencies (main data source being the Houses of Parliament's Digital Service [125], but see Chapter 4 Section 4.2 for procedure), and linked this information to the British Election Study panels (BES, [126],[127],[128], see Chapter 4 Section 4.2 for particulars) in both years to investigate campaign effects on men's and women's psychological engagement with politics using comparable survey questions. A visual overview of this empirical strategy in Chapter 4 is in Figure 4.1 on page 46.

My manipulated images mirror gender diversity at theoretically relevant numerical proportions, those that might be observable in legislatures or government cabinets. Previous empirical studies on the role model effect largely neglect this aspect, treating numerical balance as a continuous predictor variable [26],[18], where a potential problem is that the effect size would capture the impact of an 'average rate' of increase. While such approaches offer important findings, they are insenitive to any non-linearity, such as if marginal effects flatten out as women become less of a novelty. For example, looking at US House races, a recent study finds that only 'mixed-gender' (a woman challenging a man) races increased political interest, but not 'women only' races [129]. In my framework, I will concentrate on five discrete categories. The 0% or *men only* context (often serving as baseline category) mimics the proverbial 'men's game' politics [130], the status quo especially in early American investigations of descriptive representation where the majority of House races saw no women [4]. Today, UK races continue to feature around 20% or a *token woman* (see descriptive findings in Chapter 4, especially in Table 4.8). While a striking minority presence, female legislators in a token position have shown to be exceptionally active advocates of 'women-friendly' policies with good success rates in bill passage [117]. Women's growing, *'critical mass'* presence is perhaps the one that is most researched in the descriptive representation literature [131], with mixed empirical support behind hypotheses that certain proportions in numerical representation, usually around 30–35%, are necessary and sufficient conditions of improving women's policy-level or substantive representation in legislatures [132]. My picture stimuli also explores effects under 50% or a *parity* presence of men and women, the ideal target of quota policy.

I investigate two sets of hypotheses about mediation in two separate empirical chapters. Both chapters use survey experiment data from the replication round mentioned above, where mediation was investigated. Both follow a combination of approaches to detect mediation: firstly by asking the right questions after delivery of the picture treatment, and secondly by research design: by implementing specific experimental manipulation on the mediators. In Chapter 5 Study, I look at self-reported anxiety across two measures to test Hypothesis 1 that role models in political office increase women's political engagement because they alleviate anxiety stemming from stereotypes about women's competency in politics. A flowchart of this strategy is in Figure 5.2 on page 97. Similarly, Chapter 6 Study 1 observes participants' ratings of policy competency on the group level, testing Hypothesis 4 that role models in political office increase women's political engagement because they are expected to contribute to a better policy output. By contrast, Study 2 in Chapter 5, and Study 2 in Chapter 6, are crossover experiments within the original framework: After delivery of picture treatment, they introduce further negative manipulation on a sample of the treated, that is, those who viewed pictures featuring at least a token presence of women and attempt to reverse role model effects: to re-induce anxiety (see also Study 2 middle panel of Figure 5.1 on page 93), and to remove women's competency expectations. If role models exert impact through these variables, the randomly selected sample receiving negative manipulation should 'disengage' from politics at levels comparable to men only exposure. Chapter 5 also contains a third study, an exploration of the role of affect beyond anxiety.

Further explanations are provided in the individual studies including more details about the underlying concepts and what I suggest to infer from these sources. Below I sketch the three types of data sources and give an overview of procedures.

3.2 Data

In Chapter 4, I triangulate three data types to tap role model effects: (1) a small *N* lab experiment where attention is investigated with eye tracking devices; (2) a larger *N* online survey experiment where psychological engagement with politics is investigated across two data collection rounds; (3) and large *N*, observational samples from the British Election Study to investigate campaign effects, that is, change in psychological engagement with politics as a function of women's campaign activity, depending on their numerical presence. Chapter 5 and Chapter 6 use the second, 'replication' round data collection of the survey experiment, analysing the embedded mini experiments excluded from Chapter 4. I therefore discuss these procedures below.

3.2.1 Lab experiment

Sampling

In the eye tracking experiment, I recruited female undergraduate students from various colleges at the University of Exeter. Twenty minutes lab work earned a reward of £10 in

Figure 3.1: The eye tracking device.
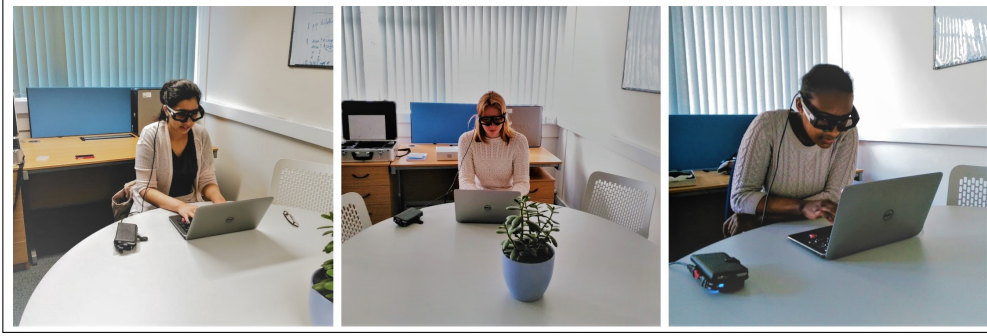
cash. The study was advertised by module conveners on email lists as well as on Facebook groups. I aimed for a sample of respondents N=30, based on the following formula, standard in test power-calculations. The dependent variables of gaze data including fixations and saccades (see later) are counted on a series of 20 pictures, the unit of analysis. Fixation *duration*, a key continuous measure of gaze being held on a particular point, is measured in milliseconds but is subject to the eye-tracker's precision in terms of sampling frequency. On a video recording obtained with 50Hz re-sampling rate, this is somewhat above 20 ms [133]. I thus determined the smallest effect size that might be detected comfortably without too much concern about measurement error should be above that, 30 milliseconds. Taking $\mu_1 = 300$ milliseconds to be gaze duration on unengaging stimuli—this would allow exactly 4 fixations across two pictures in a pair presented for 1,250 milliseconds in total[5], see later—, gaze duration of $\mu_2 = 330$ milliseconds on engaging stimuli, assuming large standard deviation under uneven attention $\sigma_{1,2} = 100$, with the critical z scores of the conventional alpha ($\alpha = .1$) and power ($\beta = .8$) levels, using the formula,

$$N = \frac{(\sigma_1^2 + \sigma_2^2) \cdot (z_{1-\alpha/2} + z_{1-\beta})^2}{(\mu_2 - \mu_1)^2} \tag{3.1}$$

I obtained the necessary sample size as $N = 137$ per group thus $N = 274$ across two groups. Measured across 20 picture trials, this comes down to a subject-level $N = 13.7$. Beyond this minimum effect of attentional bias between pictures of improved versus decreased gender balance, I aim to further detect following effects for which my calculations are less strict due to the costing of the experiment: on the one hand, Chapter 4 aims to look at the effect of the particular gender balance such as a token presence of women or a critical mass of women I thus aim to isolate a few pictures where these proportions are used. Chapter 5 Study 3 is a randomised affect study (see below), for which I need 2 treatment groups and 1 control group. Therefore, my target sample size is 30 students which may accommodate 3 group effects given potential missing data on trials such as due to unobserved pictures.

Main procedure

Wearing head-mounted eye-tracking devices, experimental subjects viewed pairs of pictures appearing simultaneously on the screen, depicting political groups with varying

men-to-women ratios. Following a brief (500 ms) fixation (**+**) in the middle of the screen, subjects were presented with a picture featuring fewer women if any at all, the 'Contrast picture', and another picture with an improved balance, the 'Role model picture'. Pairs appeared in randomised order, with 8 unique pairs appearing maximum 4 times. Each participant was presented with 25 pairs, selected randomly from this pool with repetition, the first 5 being practice trials to familiarise with the task, these are excluded from analysis. Each single exposure lasted for 1,250 ms, and was then replaced by a probe on the screen (**x**). The probe replaced either the left or the right picture, in randomised order, and participants were asked to respond to the appearance of the probe as quickly as they can by pressing the key 'E' for a left probe, or 'I' for a right probe. I programmed the task using the free trial version of Inquisit 5.0 and all other procedures of the experiment (information, consent, the exit survey, and debriefing material) in html. The experiment ran on my personal laptop computer to enable better mobility. Gaze data was stored on the unit attached to the eye tracker, consent, response latency and survey data was stored on the computer.

I selected the pictures with the following procedure. To improve external validity, the aim was to select real-life and credible campaign pictures of political groups. To my knowledge, no such tested pool of pictures exist for experimental purposes. To control for exogenous effects on attention and response strategy, I first needed to make sure that the figures are not well known politicians, that no party labels and connected identification (flags, other logos) remain in the pictures. Knowing it would be difficult to control for other features including ethnicity, age, dressing, I chose to induce *more* variation in these features and aimed for a large pool of diverse figures.

Following the UK General Elections of 2015, I pooled together data on all running candidates (see Section 3.2.3.) and selected those that achieved less than 5 percent of the vote share in their respective constituencies, minimising the probability that they may be recognised. Since my study sample currently resides in Exeter, Devon, I excluded candidates running in all Devon constituencies. After this round of selection, I turned to publicly available, tweeted images of these candidates during the GE2015 campaign—I obtained twitter handle data from the Media in Context project [134]—, with the aim of selecting those that have both gender represented in them, as well as some where politicians are of varying ethnicity.[6] On all images, participants are facing the camera with facial characteristics indicating positive mood.

To compile the image pairs, I individually cropped women out of most pictures, so that I could pair men only/mostly men groups with the same original picture but with an improved gender balance in them. I changed some but not all of these image pairs on a grey scale to control for attention bias possibly stemming from vibrant hair and dress colours. The pictures are paired according to the following percentage presence of women: four pairs featuring 0% (men only) Contrast picture and a near 20% or presence Role model picture; another pair contrasting 0% with women's 30% or critical mass presence in the Role model picture, and another contrasting 0% with 50% or parity in the Role model

Figure 3.2: Picture contrasts—lab experiment.

Each participant completed a series of 20 trials (2 examples below), presented in random order for 1,250 milliseconds, one of each replaced by a probe in each trial. Original pool presented in Figure 3.3

Trial #1/20 (example)

Screen 1: Fixation screen, 500 ms — Screen 2: Stimuli, 1,250 ms — Screen 3: Probe screen, 500 ms

Trial #2/20 (example)

Screen 1: Fixation screen, 500 ms — Screen 2: Stimuli, 1,250 ms — Screen 3: Probe screen, 500 ms
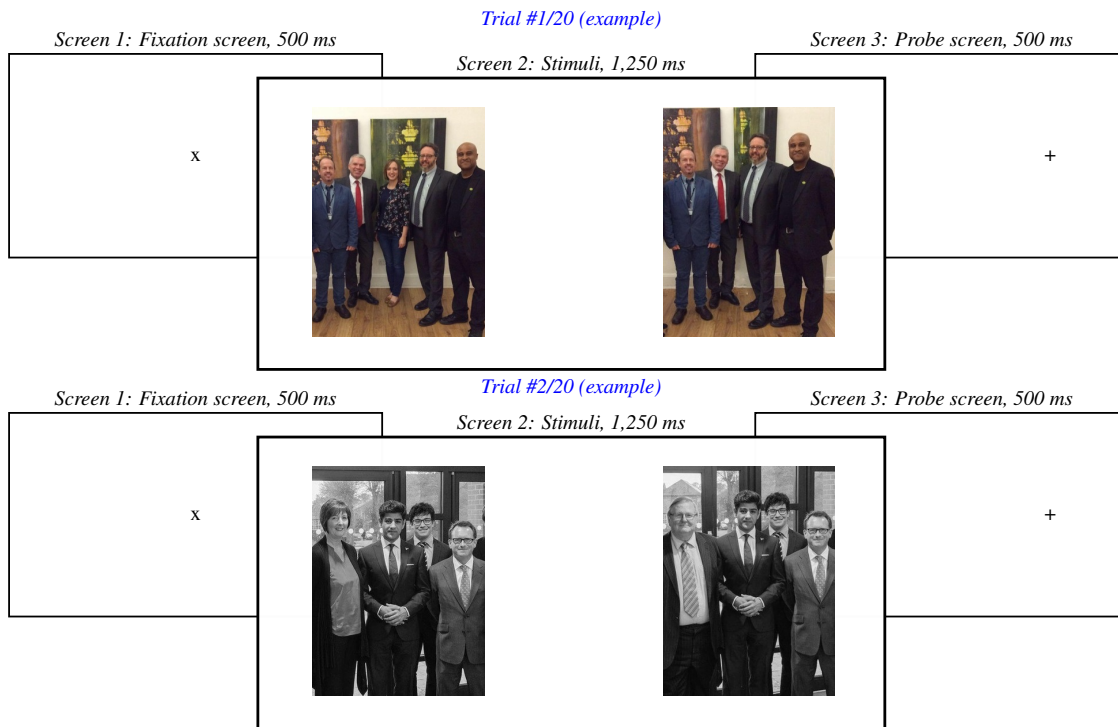
picture. Two Contrast images had women in them: a token and a critical mass picture, both paired with women majority pictures.

Figure 3.2 shows example picture pairs within the flow of the trials in the dot probe task. The full picture pool from which the contrasting pairs are selected are shown in Figure 3.2. My pictures being previously untested in previous dot probe tasks, my exposure time of 1,250 ms is exploratory. Simpler picture stimuli (such as one face per picture) are typically presented for a shorter period of time but with more complex stimuli, researchers could detect attentional bias typically at 1,000–1,250 ms presentation times [135],[136]. I reasoned that 1,250 ms would be enough minimal exposure time for basic exploration and a few brief fixations, and that longer exposure would risk the possibility of conscious processing.

My dot-probe task script itself logs response latency to probes, along with the information if it replaced the Contrast or the Role model picture. The simple differential of these average latencies constitutes the estimate of attentional bias. In Chapter 4, I will use dot-probe measures parallel with the eye-tracking measures, treating them as alternative approaches and looking for consistent patterns.

After completion of the task, participants carried out an additional task measuring implicit associations regarding gender and leadership, filled out an exit survey, and got debriefed. Completion of these additional tasks was not recorded by the eye trackers to save on video size. None of the additional data is analysed in my thesis Chapters, being out side the scope of my studies.

Emotion induction

In Chapter 4, I analyse if attention is split evenly or unevenly across Role model pictures and Contrast pictures. In Chapter 5 Study 3, I analyse these patterns conditioned upon emotion, guided by hypotheses H2 and H3 (See Section 2.1.3. and Chapter appendix 2.5). Following a relatively unobtrusive method in emotion research [137],[138],[139], I randomly assign emotions, namely, of approach (anger), and of avoidance (sadness), as introduced in the previous chapter, to subjects by making one, rather than the other, emotion salient before task completion. I am aiming at the induction of no relevant emotions for the control group.

Firstly, they are presented with short paragraphs of text about women's experience with everyday sexism. These are user-generated comments from an online community called the Everyday Sexism Project, www.everydaysexism.com. Using dictionary-based, computer-aided text analysis with the LIWC software [140],[141], a method I present more in detail especially in Chapter 5 Section 5.2.3, I present predominantly 'angry' comments in the anger group, and select those that have an explicit component of action in it. For example,

> "[...] The most tragic with [this] story isn't harassment itself: it was that none of them in the team seemed like they were going to tell him off, using the 'ignore and he'll go away'-tactic. The only thing one of the senior players said about it, was mentioning to one of his team mates they better keep an eye of him. His team mate answered they were used to him behaving like that. Well, I am not. *And it is a must to fight their ignorance. No one else will do this for me.*" (highlight added)

In the sadness or helplessness group, I manipulate content to remove the action component from the story, as well as include other segments within the same stories flagged up by LIWC as 'anxious' or 'negative affect' but not 'angry'.

> "[...] The most tragic with [this] story isn't harassment itself: it was that none of us in the team seemed like we were going to tell him off, using the 'ignore and he'll go away'-tactic. The only thing one of the senior players said about it, was mentioning to one of his team mates they better keep an eye of him. His team mate answered they were used to him behaving like that. *I shudder at the thought what that man is capable of when he's drunk and there's no one of his friends around to keep him in check.*" (highlight added)

By contrast, the control group reads user comments, selected from Amazon.co.uk product reviews.

> "Today is the fourth day of ownership of my pen, and I have to say I'm starting to treat it like an old friend. I would happily recommend this pen to anyone who is planning on writing on paper."

Secondly, once finished with the reading task, they are asked to carry out a short writing task. In the emotion groups, they are asked to write 'a sentence or two about something—a personal story, or [participants'] reflection of what [they] read—that makes [participant] *angry about sexism/sad of being subjected to sexism*, let that be [their] own experience or some else's who [they] have heard of'. The control group receives the neutral task of writing a review of recently bought products. Participant's answered are analysed to check treatment compliance at the relevant sections in Chapter 5.

Data overview

Dependent measures are used to test the hypothesis that Role model pictures with an improved gender balance are more engaging than Contrast pictures, and measured across: (Models 1.0–1.2) response latency on the dot-probe task, (Models 2.0–2.2) fixation—gaze held at fixation point—duration on pictures, (Models 3.0–3.2) fixation count on pictures, (Models 4.0–4.2) duration of saccades—rapid movement between two fixation points—, and (Models 5.0–5.2) count of saccades. I also situate these measures in the context of attention research in Chapter 4.1.1.

The dot probe paradigm builds on a simple mechanism, namely that fast probe detection, all else being equal, follows probes that appear in *areas that were 'attended'* by the test subject, and *slow detection* follows probes that appear in *'unattended' areas* [75]. Attentional bias is in turn defined as the systematic differential in response latency—speed in probe detection, measured in milliseconds—for probes replacing more engaging stimuli minus response latency for probes replacing less engaging visual stimuli.

More direct inference about attention may come from gaze data. While I am only able to detect one dot-probe response after each trial, gaze might have a distribution across Contrast pictures and Role model pictures within the same trial. In terms of fixation, inference works as follows: increased duration of fixations, measured in milliseconds, as well as a greater number of fixations within the same (type of) picture should indicate greater engagement with these locations. In terms of saccades, my analysis is exploratory: while saccades are involuntary eye movements, an increased frequency of saccades within the same picture may signal that the eye re-focused multiple times to observe multiple locations. There are no explicit hypotheses about saccadic duration.

Key independent measure is picture type, being either Contrast or the Role model picture. To account for numerical balance, another independent variable will be the proportion of women in the Role model picture, as outlined above. In the additional Study 3 in Chapter 5, another key grouping factor is emotion treatment group membership. Measures taken but not analysed within the scope of my thesis are performance on a 'custom-built' gender–leadership Implicit Association Task, and the exit survey exploring political engagement as well as political values that is identical to that carried out (and analysed) in the large N online experiment.[7]

3.2.2 Survey experiment

My survey experiment project comprises two data collection rounds: an *original study* conducted in May 2015, and a *replication study* conducted in June 2017; all data collected via crowd sourcing platforms 'Prolific Academic' and 'Crowdflower'. These are large pools of paid survey participants with the advantage that UK-based samples may be targeted with a wide demographic distribution, emerging as viable alternatives for 'old' and overly professionalised survey worker platforms, and tested for data quality [142]. In the first round, I used both platforms with respondents pooled across samples. I correct for unobserved differences across sample respondents via hierarchical modelling, see Section 4.2 Equation 4.5. In the replication round, I only used Prolific Academic which proved to be simpler in terms of data management, and the only platform that allows me to recruit an even split of men and women.

The two rounds of data collection are identical in procedures to the end of detecting role model effects on political engagement (Chapter 4), the setup I called 'blank causation', I thus present results pooled across both rounds of data collection. Again, hierarchical modelling is useful to account for unobserved differences, this time across data collection rounds, but I do present descriptive results separately as well. Mediation (Chapter 5 and Chapter 6) is tested exclusively in the replication round where relevant measures and alternative procedures are introduced on top of the original procedure: this is called a crossover design in natural science, where a sequence of treatments are introduced one after another. Alternative procedures that might interact with the original procedure were administered on isolated subsamples only, and those respondents are evidently excluded from the blank causation study.

Sampling

Using the identical formula for sample size calculation, presented above in Equation 3.1, I determined that on a *z* score-type, scaled-centred continuous dependent variable, where the mean is $\mu_1 = 0$ and $\sigma_1 = 1$, a substantively significant effect of one standard deviation, $\mu_2 = 1$, $\alpha = .05, \beta = .8$ would require 16 participants per group. I am increasing this target having the following considerations in mind: Firstly, in my research design, as explained below, I am aiming to detect effects not across two but 4 treatment groups (exposure to men only, a token, critical mass, or parity presence of women in picture stimuli), allowing, in fact often explicitly hypothesising, that treatment effects differ across the two gender, thus this number goes up to $4 \cdot 2 \cdot 16 = 128$. Secondly, there is no straightforward rule as to how this might be modified in a hierarchical model using random effects, teasing out variance components across crowd sourcing platforms and data collection rounds after sample pooling. Random effects terms increase the posterior variance of estimates (the effect sizes of interest), thus larger sample sizes are useful to confidently detect treatment effects. Thirdly, while I treat many of my scalar dependent variables as continuous measures for better comparability across data collection rounds and with the large N probability sample, that one standard deviation difference is fairly difficult to detect with few

scale points.

I thus concluded that I needed sample sizes near 300 for both data collection rounds when comparing effects of 'blank' causation. In the replication round, this number increases to 750 to accommodate further manipulation used in the mediation studies, using the following logic: if 3/4 of the sample receives picture stimuli featuring at least a token woman politician (see next section), that is, 225 out of 300, I need that an equally sized group receives negative manipulation on the anxiety mediator (Study 2 Chapter 5), and that an equal number of participants receives negative manipulation on the policy output mediator (Study 2 Chapter 6), in order to make the appropriate comparisons, totalling $300 + 225 + 225 = 750$. 'Treated' subjects in the 300 pool serve as control group for the randomised mediator groups. Figure 3.4 on page 40 gives an overview of these groups.

Participants received incentive in form of a pro-rata payment of the UK minimum wage, £7.50, which is £1 for 8 minutes of survey work.

Main procedure

The main procedure applies to both rounds of data collection. Participants were informed that that the survey is about political leadership evaluation. Randomly assigned into four groups, an equal split of men and women viewed a series of five pictures featuring 2015 political campaign groups (information given explicitly in instruction). Gender balance in picture stimuli varies across groups but not within, thus any given participant was only exposed to men only politicians, or a group featuring a token woman or 20%, a critical mass of women or 30%, or a parity of men and women or 50%. I introduced this picture pool in detail above in 'Main procedure' Section 3.2.1, comprising tweeted images of less known campaigning politicians with a popular vote share below 5%. A selection of images is directly comparable across groups, that is, I individually cropped out women in varying proportions and assigned them in different treatment groups. Some groups needed to be filled out with new pictures as well, these do not have a counter part in other groups. Figure 3.3 shows all pictures used in the treatment groups.

Survey participants did not merely observe these pictures: they rated them on policy areas. The question, under each picture, reads 'How well do you think this group would manage [finance/healthcare] policy?', and I analyse the response scales themselves in Chapter 6 Study 1.

After completion of the viewing and rating task, participants in the first round of data collection moved on to complete a survey of psychological engagement with politics. In the replication round, survey participants first responded to questions about state emotion, and completed a short reading and writing task. All participants viewing men only pictures, as well as a third of all other treatment groups completed a neutral task of writing about recently bought items. While serving as control groups in the randomised mediation studies, see 'Alternative procedures' below, I am also confident to use the answers of these participants as replication data for the original study on role model effects. Both rounds ended with debriefing, giving information about the aims of my thesis.

Figure 3.3: Picture pool—survey experiment.

Each treatment group viewed a series of 5 pictures, rating them on 'healthcare' or 'finance' policy management (randomised).

Figure 3.4: Overview of sample management
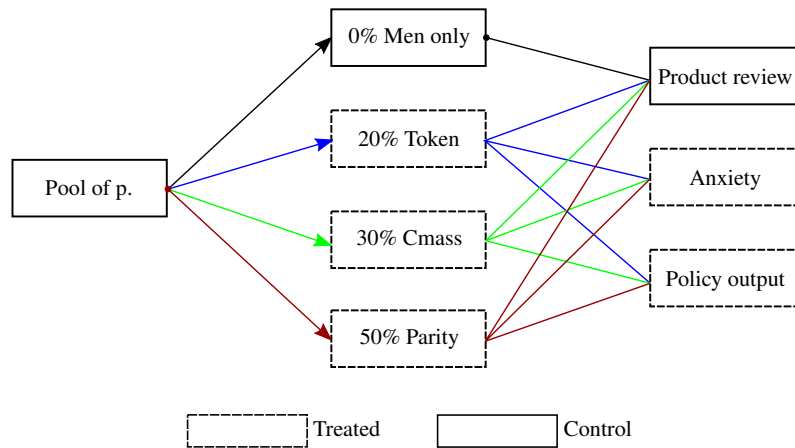
Alternative procedures for randomised mediation studies

Similarly to the emotion induction procedure in my lab experiment, participants completed a combination of tasks in the replication round of data collection. In Chapter 2 I outlined two ways role model effects may make sense on the individual level namely that they alleviate stereotype threat and predominantly anxiety, and that they signal competency in 'women friendly' policies. My aim in the randomised mediation studies, using the principle of the crossover study design, is to explicitly *remove* or decrease these expectations, that is, to observe if role model effects diminish on psychological engagement with politics, once mediators are removed. On anxiety as a mediator, this hypothesis is shown in Table 5.7 on page 106. To my knowledge, there are no comparable approaches in the field of political representation but they are extremely rare in any sub-field of social science as well.

The anxiety group's main task is to 'take a moment to think about and briefly describe 3-5 things that do make [them] anxious', knowing that they did not 'need to take too long on this one. Few words or a sentence would suffice'. Once done, they moved on to another writing task in which they were asked to 'describe in more detail the one situation that makes [them] the most anxious so that someone reading will become anxious'. Simultaneously, the policy output group was asked to 'take a moment to think about the following two findings taken from a recent expert study about women in politics in the UK', which said

> The notion of women's feminizing of politics is controversial [...] implying that as a group, they inherently share a set of policy preferences or ideology. Women are not a monolithic entity with a collective set of interests and beliefs. One could cite many examples of female political figures who represent the antithesis of what many regard as "women's interests" (e.g., Margaret Thatcher). Representation [...] is not a guarantee that women representatives will act for women. [...] In finance debates, findings show that the most significant [factors in] debate participation are relevant committee

membership, Conservative party [membership], and seniority, holding other explanatory variables at constant. Sex was not significant.

which was meant to motivate thinking about women textnot pursuing a woman-friendly policy. Once done, these participants were asked 'How would you describe the main finding of the authors? Should you need to read the text again, navigate back below'. The control group received emotion-neutral task of listing recently bought grocery items, then describing one in detail.

Data overview

The survey experiment questionnaires are attached in the appendix. I aimed to have comparable items to observational studies. While in my thesis, I chose the British Election Study as benchmark, it was not the only option considered, thus across the two data collection waves, I experimented with different response scales to measure psychological engagement with politics. To make them comparable between my two data collection rounds, I need to transform these scales to arrive at *z*-score type, 0-centred and standard deviation unit measures. I give more details on the original distributions separately as well in Chapter 4. In Chapters 5 and 6, I am using data only from the second round of surveying, thus the original scales may be kept and modelled more appropriately to measurement level.

I review the key dependent variables, measuring psychological engagement with politics, at the relevant sections namely political interest in Section 4.2.1.; political knowledge including test questions in Section 4.2.2., political efficacy in Section 4.2.3., and trust in MPs in Section 4.2.4. Some of these are single questions ('How interested would you say you are in politics?'), and some of them assessed on multiple questions such as a political knowledge quiz, confidence in political knowledge assessed after each test question, or political efficacy. My analytical strategy is to use hierarchical modelling wherever possible, therefore as dependent variables, these multi-item answers, such as self-assessed confidence after each test question, are nested within latent concepts, namely confidence in political knowledge, as well as within individuals (See also Equation 4.6 in Chapter 4 Section 4.2).

Key independent variable is treatment group, exposure to 0% or men only picture stimuli, to 20% or token stimuli, 30% or critical mass stimuli, or 50% or parity stimuli; as well as its interaction with respondent gender. Model building is explained step by step, using model equations to demonstrate which group means are derived, first in Chapter 4.

3.2.3 British Election Study (BES) Internet Panels

I am using the BES panels to estimate campaign effects on psychological engagement with politics, rather than looking at them cross-sectionally, to see if exposure to women politicians in campaign *increased* women's political engagement, in a within-individual design. Both the 2010 and 2015 election study panels have a pre-election wave (if multiple, I chose the closest to the campaign period), and a post-election wave. The 2015

study internet panel has a number of campaign waves as well, these have diffuse timing and I have no information about the distribution of women's campaign activity during this period, I thus discarded these waves.

Data overview

Dependent variables are scaled-centered changes in political interest, political efficacy, and trust. In terms of political knowledge, on a small set of three questions about foreign leaders, I am able to determine if participants learned about politics through the campaign. I give more details about these variables in the relevant Sections 4.2.1 through 4.2.4 per dependent variable.

I link the BES data, where respondents' parliamentary constituency identifier is available, with Electoral Commission data on the candidates running in the respective constituencies[8]. While candidate names are readily available for these election years, candidate gender had missing data for 2010. Wherever the same person, 2015 candidate gender is imputed on these places. Further missing data was filled in using crowd-sourced information site Democracy Hub, the 'babynames' data set relying on historical US Social Security Number Application data to predict probable gender given first name, and finally, hand coding residuals. I thus obtain 'treatment' comparable to the survey experiment varying from 0% men only constituencies but surpassing parity and including women majority constituencies as well. The distribution of gender balance across parliamentary constituencies is shown in Table 4.8 on page 59.

Similarly to the survey experiment, another key independent variable is respondent gender. Finally, I need to include a minimal host of covariates, this being an observational study, on respondents' characteristics influencing political engagement: age and education, and on constituencies' characteristics influencing both political engagement but perhaps also the distribution of 'treatment': seat marginality. I discuss these measures within the voting behaviour literature in Chapter 4 Section 4.2.

This chapter was meant to give an overview of key methods, procedures, and variables collected. I give more details on measurement, question wording, missing data management, and analytical strategy in the individual studies.

Chapter 4

Measuring role model effects:
Experimental and observational approaches

> If women inhabit a less political world than men do—one that provides less exposure to informal political chat and other politicizing cues and offers fewer relevant political role models—then women are likely to infer that politics, like football, is not for them. [17, p.8]

> *In this chapter, I report observational and experimental evidence behind two types of role model effects: of tokenism, where attention and political interest-related variables are impacted by women's striking minority presence in politics; and of linear role model effects, where a gradually increasing proportion of female politicians gradually increases women's self-efficacy and confidence in politics.*

Existing empirical evidence behind the role model effect specifies if and when women politicians engage fellow women in politics. Results converge on a positive impact on *political knowledge*, as women are more likely to recall political information including names of representatives if living in areas where high visibility offices such a Senate seat or an executive office are held or contested by women [7],[23],[17],[24],[25],[143]. Similarly, evidence is robust that women and especially young girls are likely to engage in more *political discussion* with friends and family across a range of contexts with more women in politics [19],[18],[25],[21]. Studies seeking to isolate effects on *political interest* and *political efficacy* offer mixed evidence [24],[26],[129], a substantial stream of American scholarship stressing partisan appeal engages people rather than weak gender identification [28],[29].

Yet a lot of this evidence relies on observational data, making it difficult to disentangle whether contexts where role model effects are detected are simply cultures where women are more politically engaged. To strengthen claims of causality, a recent panel study provides confirmatory analysis that novel women in campaign do stir more political discussion among young women but does not measure political engagement beyond that [21]. Similarly, an online experiment finds some effects on mobilisation (voting propen-

sities) but fails to uncover an effect of women major party candidates—Republican or Democratic—on campaign interest and political efficacy, casting doubt on the role model hypothesis altogether [27]. By contrast, a very high external validity natural experiment in India [61] reports that one of its states' gender quota policies, whereby a randomised third of all village councillor positions were contestable by women only, increased young girls' career ambition in numerous fields including leadership roles. While similar experimental evidence exists that (political) role models may close the achievement gap in careers and academia such as evidence of an 'Obama effect' [56], or a 'Hillary Clinton effect' [58],[59] (see especially Section 2.1.2), effects remain to be investigated in terms of psychological engagement with politics.

In this chapter, I set out to contribute to this empirical puzzle by considering a wide range of measurement options on the one hand, and by combining experimental and observational evidence, on the other. As point of departure, I study whether women's *attentional engagement* with a political group varies as a function of its gender composition. I designed a lab experiment in which responses to a cognitive test as well as eye-tracking measures were collected in response to visual stimuli depicting political groups, allowing me to look at how the varying degrees of gender balance in groups attracted women participants' attention. Parallel with this study, I implemented an online survey experiment where a larger sample of men and women, following exposure to the same pictures and randomised across gender balance, filled out a standard survey of political engagement. In this study, I explore if women's *political engagement* increases following improvements in gender balance, and whether patterns of political engagement follow patterns of attentional engagement found in the lab experiment. Finally, I 'recreate' some of the gender balance setups from the picture stimuli by identifying the proportion of women candidates running in the 2010 and 2015 United Kingdom General Elections in each parliamentary constituency, and link this information to the British Election Study internet panels to investigate if citizens' reported levels of political engagement *increased* during the two campaign periods, conditioned upon gender balance in their constituency's race. An overview of this empirical strategy is displayed in Table 4.1.

Situated within the larger empirical project of my thesis, the aim of this chapter is mainly to gather evidence of cause and effect or 'blind causation', pooling together a large number of relevant variables. This prepares the scope for the subsequent chapters tackling more difficult questions of mediation: channels explaining why exactly a greater presence of women in politics should engage fellow women, namely by alleviating stereotype threat and implicating affective processes such as anxiety (Chapter 5), and by influencing expectations about greater efficacy in implementing policy including 'women friendly' policies (Chapter 6). It is entirely possible, however, that mediation will be open for testing for some variables where I find at least minimal experimental evidence of an effect but not for others.

I discuss my results by key dependent variables, split by level of abstraction. Section 4.1. reviews psychophysiology evidence to infer attentional engagement, while Section

4.2. reviews survey evidence to infer more explicitly on political engagement. Latter is a host of measures including political interest in Section 4.2.1., political knowledge in Section 4.2.2., political efficacy in Section 4.2.3., and trust in MPs in Section 4.2.4.

Across these sections, common is the analytical strategy in which some form of a Bayesian hierarchical or mixed-effects model is fitted with a comparable host of predictor variables to tap role model effects. While this may complicate estimation of treatment effects beyond comparing group averages, mixed-effects modelling is necessary to account for the hierarchical structure of the data in which observations are not always independent from each other. In the lab experiment models, hierarchical structure describes how test responses and gaze data on picture stimuli are nested, foremost, within experimental subjects, suggesting that two measures on gaze durations would be related to each other taken on the same test subject. In the models using survey data, many dependent variables such as test performance or political efficacy are *latent* meaning they are measured across multiple 'response' items or test questions but tapping the same underlying construct, in this case those items are nested within latent constructs. Finally, in the BES models, 'treatment' varies on the parliamentary constituency, not the individual level, thus survey respondents are nested within their respective constituencies as well. Bayesian hierarchical models are suitable to make these structures explicit while making the estimation of fairly difficult models with interaction effects across levels possible—those models may prove difficult or impossible to estimate in the classical statistical models supporting frequentism. In the following sections, the term *random effect* is reserved to denote variables that describe hierarchical structure (e.g. subject ID, test item, constituency)

In terms of comparable predictor variables, in each section 'treatment' is the proportion of women that study participants are presented with, either on picture stimuli or in real-life campaigns. To allow for varying or non-linear treatment effects, I do not treat this as a continuous variable but look at theoretically founded, key discrete values of women's presence. The 0% or 'men only' context (often serving as baseline category) mimics the proverbial 'men's game' politics [130], the status quo especially in early American investigations of descriptive representation where the majority of House races saw no women [4]. Today, UK races continue to feature around 20% or a token woman (see descriptive findings, especially in Table 4.8). While striking minority presence, female legislators in a token position have shown to be exceptionally active advocates of 'women-friendly' policies with good success rates in bill passage [117]. Women's growing, 'critical mass' presence is perhaps the one that is most researched in the descriptive representation literature [131], with mixed empirical support behind hypotheses that certain proportions in numerical representation, usually around 30–35%, are necessary and sufficient conditions of improving women's policy-level or substantive representation in legislatures [132]. Finally, I explore the effect of 50% or parity presence, the ideal target of quota policy, and in the BES models, women's majority presence, observed around 60%.

Participant gender is interacted with treatment in the survey experiment and BES panel studies so that unique, women-specific role model effects can be isolated, men

45

Table 4.1: Chapter 4—Empirical strategy

Study type	Target concept	Operationalisation
Lab experiment		
	Attentional engagement	
		Attentional bias 1.
		(Dot-probe task response)
		Attentional bias 2.
		(Eye-tracking measures)
Survey experiment		
	Psychological engagement with politics	
		Political interest
		Political knowledge 1.
		(Subjective self-report)
		Political knowledge 2.
		(Test performance)
		Political knowledge 3.
		(Test answer confidence)
		Political efficacy
		Trust in politicians
Observational study		
(British Election Study, pre- and post-campaign waves, 2010 & 2015)		
	Psychological engagement with politics	
		Change in election interest
		Change in political knowledge 1.
		(Subjective self-report)
		Change in political knowledge 2.
		(Test performance)
		Change in political efficacy
		Change in trust in politicians

being the framework of comparison. In the lab experiment using eye-trackers, due to its higher cost and exploratory nature, I recruited women participants only thus this element of comparison is entirely missing.

Finally, as appropriate given the data type at hand, a host of covariates is included. In the observational study, these are age and education on the individual level, and winning majority on the constituency level—a readily available indicator of party competition, which may (though imperfectly) indicate if a seat is safe or marginal, often with consequences on political interest [144]. Moreover, historically and especially before the 1997 General Elections, UK major parties have tended to place women in unfavourable seats challenging incumbents in a safe seat [145]. Therefore, there is a possibility that 'treatment' in the BES is non-randomly distributed across constituencies but depend on winning majority. In the survey experiment, age and education factors are assumed to be randomly distributed across treatment groups (an issue to be investigated), but I do include data collection platform and data collection year (an original round of data collection in 2015 and a replication round in 2017, as random effects allowing me to pool data gather across different 'time or place' but accounting for unobserved differences (variances) at the same time. More explanation on random effects are provided below near the model equations, but see also Chapter 3 Section 3.2.2. In the lab experiment models featuring eye-tracking generated data, I control for 'probe location', as elaborated in the next sections, to discount non-random measurement error: imprecise timing

through data extraction whereby gaze data remains collected after the pictures disappear from the computer screen for a very brief time.

The next sections review main results with minimal speculation about their wider significance. I present statistical models expanded gradually, starting from uninformative or 'null' models, then including treatment and/or gender effects estimating the average treatment effect (ATE), and lastly including covariates or any further manipulation deemed necessary. Similar approaches of model expansion are common in hierarchical modelling allowing researchers to identify variances on multiple levels of nesting (e.g. is there any variation on the individual level to be explained? is there a variability on the constituency level?), which I will often reference as well. Occasionally, movements in variance components may be used as partial substitutes for the classical 'model fit' statistic such as the R^2 which is missing in extensions of linear models. In the first examples of dependent measures ('Attentional bias 1.'), this approach is explained step by step but will later be omitted, concentrating on the key effect sizes.

Substantive interpretation is left to the Section Conclusions 4.1.3. for Attentional engagement, and 4.2.5. for Political engagement, while Section 4.3 draws more general conclusions concentrating on common patterns across these studies. A wider significance of my findings is discussed in my thesis Conclusions chapters.

4.1 Attentional engagement

If somewhat indirectly, attention has been in focus of political behaviour studies, especially on information seeking, selectivity, and learning [146]; and more recently, attentional bias has been measured to describe how Conservatives and Liberals react differently to simple valenced stimuli [147] as well as complex emotional tone in campaign [148]. The current study, however, is novel in the sense that it measures attention more directly to infer political engagement. I use basic physiological evidence of attention, as in engagement with visual stimuli from the political world, as early indication of political interest, bearing in mind consequences on the acquisition of knowledge of current affairs via learning.

Attentional engagement might thus viewed as micro-level evidence of political engagement in a paradigm where competing political stimuli are available to collect information from. In this section, I describe the results of a small, $N = 28$ lab experiment where an incentivised convenience sample of women undergraduate students completed a cognitive task called the dot probe test while wearing head-mounted eye-tracker devices, collecting large amounts of data per millisecond of gaze location, duration, and saccadic movements. The task itself was to view 25 pairs of images, all depicting real political groups. The pairs featured very similar images except that the women in them were cropped out in varying proportions. One of each pair featured a group whose gender balance was improved (with more women in them), compared to the other with fewer or no women at all. Throughout this analysis, I refer to the former set of pictures as the role model pictures, and latter as their Contrast pictures, within each pairing. Both may have

appeared on the left or the right. Each pair was replaced by a dot 'probe' on either side, and participants were required to identify which side, 'as quickly as possible ' (Figure 3.2 in Chapter 3).

Ultimately, the aim is to compare attention to images of improved gender balance with those of decreased gender balance. Important detail is that the unique pairs (8 pairs, reappearing in random order and 20 times in total) were chosen to introduce *more* variation in the details of the pictures: women appeared in various positions in the frames, with varying details in appearance, including race. The idea behind this is that an average treatment effect, if detected, would more likely be due to gender balance, rather than the varying features. Given the number of trials, I chose to introduce most variation in the context where one woman is surrounded by multiple men (up to 4 picture pairs, in later sections labelled as the 'token condition') but I did not vary positions and appearance to the same extent in other numerical representation setups such as a 30% critical mass presence or complete gender parity.

4.1.1   Attentional bias 1.

Dot probe paradigm

The dot probe paradigm itself has been used to infer attentional bias [149][75] building on a fairly simple mechanism, namely that *fast probe detection* (the probe being a black dot on the computer screen), typically executed by pressing a button on the keyboard, follows probes that appear in *areas that were 'attended'* by the test subject, and *slow detection* follows probes that appear in *'unattended' areas* [75]. Attentional bias is in turn defined as the systematic differential in response latency—speed in probe detection—for probes replacing more engaging stimuli minus response latency for probes replacing less engaging visual stimuli. While the eye-tracker data is more useful to get at attention directly, for its availability, and to produce a larger number of measures for cross-validation, I first present attentional engagement patterns based on the dot-probe task performance.

The basic hypothesis is that women would pay more attention to groups of politicians with more women in them (the Role model pictures), as opposed to more stereotypical, male dominated groups (the Contrast pictures). The study is largely exploratory of attentional engagement at the specific values of numerical balance. While most picture pairs contrasted a single woman's token presence (20%) to men only groups, one or two pictures presented women in other proportions, as listed in Table 4.2.

The dependent variable is the latency, measured in milliseconds, of the submission of a correct response, that is, correctly identifying whether the probe appeared on the left or on the right hand side[9]. Across 28 subjects, responses were submitted on $28 \cdot 20 = 560$ image pair trials[10], the vast majority of which, 513 being correct.[11] On average, probes were detected in about a third of a second. A large portion of variation in response latency should be explained by individual-level variation in vigilance or skills (i.e. some people would just simply be faster to respond than others), or in terms of how individuals might get better at task execution as they advance through the trials. A 'null' model, Model 1.0,

Table 4.2: Descriptive statistics: Measures of attentional engagement

Pairing #	Prop. women	N	M ws resp. latency (dot probe)	M ws duration of fixations	M ws count of fixations	M ws duration of saccades	M ws count of saccades
1-Role model	0.20	3	368.69	258.26	12	782	5
1-Contrast	0.00	3	420.28	224.04	10	516	4
2-Role model	0.25	1	390.50	270.21	4	299	2
2-Contrast	0.00	1	387.35	201.81	3	133	1
3-Role model	0.33	1	376.70	222.01	3	100	1
3-Contrast	0.00	1	401.48	216.45	3	233	1
4-Role model	0.50	1	395.86	234.66	4	333	2
4-Contrast	0.00	1	393.43	239.51	4	266	2
5-Role model	0.67	1	428.33	258.38	5	250	2
5-Contrast	0.33	1	393.76	241.71	3	67	1
6-Role model	0.75	1	390.31	204.72	2	33	0
6-Contrast	0.25	1	396.20	225.21	2	33	0

'ws' *stands for within-subject statistic;* 'M' *for mean duration or median count.* Inference*: (a) dot-probe: shorter response time is suggestive of attentional bias across pairings; (b) gaze data: greater duration & counts may mean attention bias.*

$$Y_{latency} = \beta_0 + u_{0t}^{trial} + u_{0s}^{subject} + e \qquad (4.1)$$

where the $u_{0.}$ terms are 'random effects' reflecting where nesting of responses is to be expected[12], decomposes this latency variable into a grand mean of response latency (intercept), and a host of grouped means: for each of the 20 test trials, and for each of the 28 participants. Using this information, I calculated that subject-level variation explains up to 200 milliseconds differential in average response times (which is estimated to be about 33-36% of the total variance, see Table 4.3), but that trial-level variance averages out to zero, explaining very little of response latency.

Next, an extended model, Model 1.1 explores what room is left for other, substantively meaningful effects, on top of trial and individual-level nesting—whether responses took longer depending on gender balance in the picture primes. Specifically, two variables explore (a) whether responses took longer if probe replaced the 'Contrast picture' with more men, or the 'Role model picture' with more women; and (b) what was the specific numerical balance in the Role model pictures. Table 4.2 shows the observed latency across these categories, suggesting that attention was uneven under two extremes: women's token minority at 20% and Women's majority at 67%. In the 20% versus Contrast setting, Role model pictures received more attention but this differential turned into the opposite direction in the 67% picture versus Contrast. In the formal model,

$$Y_{latency} = \beta_0 + u_{0t}^{trial} + u_{0s}^{subject} + \beta_1 X_t^{pic.\ type} \cdot \beta_2 X_t^{\%\ balance} + e \qquad (4.2)$$

the multiplicative term teases out response latency within these categories, resulting in 2 (picture type either Contrast or Role model) · 4 ($\sim$20% or $\sim$30% or 50% or >60% balance n the Role model picture) = 8 point predictions. Looking at the marginal effects (the β coefficients in Table 4.3), Model 1.1 finds weak evidence of a token effect (see β of the non-interacted Contrast picture) to begin with. Its large posterior variance—90

or 95% credible intervals both include zero—is possibly due to my lumping 20% and 25% pictures together, but the descriptive Table 4.2 shows those 25% pairs received even attention. Moving on to increasing percentage, effect sizes diminish. There is sufficient evidence, however, behind a reverse women majority effect (>60%. x Role model picture) suggesting that as men become minority, attention is directed to them.

Finally, in Model 1.2, I excluded responses to the 25% picture to investigate if its inclusion was responsible for the token result, and in which case the 20% token effect does remain positive within 90% credible interval (but note that Table shows more conservative 95% CI). In this model, the token effect is the distance between the Intercept and the Contrast picture estimates, suggesting that it is about a third standard deviation, 33 milliseconds. The predicted response latencies, transformed from standard scores to milliseconds, are visualised in Figure 4.1.

Predicted response latency

95% distribution of posterior predictions
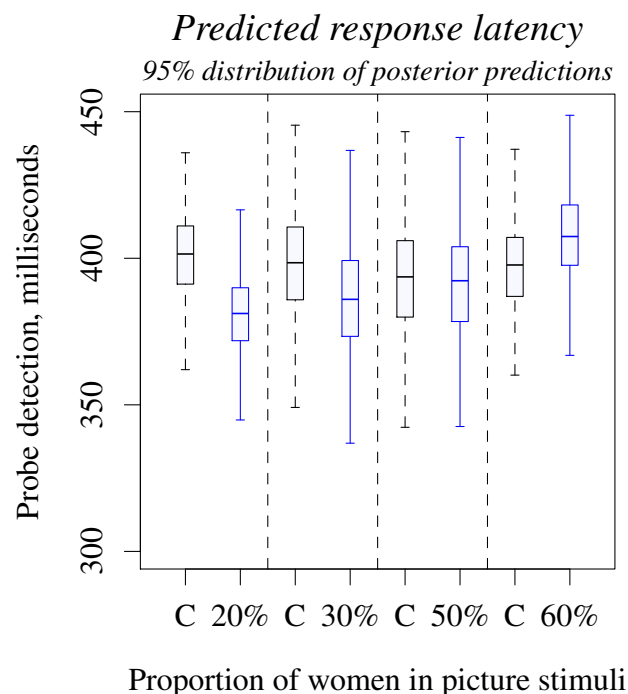
Note: C stands for 'Contrast image' in pair featuring less or no women. Faster probe detection may indicate attentional bias.

Figure 4.1: Attentional bias across picture contrasts, Model 1.2

The dot-probe task thus shows only weak evidence that gender-balanced stimuli receive more attention, that attention is even under other contrast options but stronger evidence that men's novel token position against a woman majority receive attention surplus. The eye tracking evidence in the next section is partially used as confirmatory analysis of these patterns.

4.1.2 Attentional bias 2.

Eye-tracking data

Response latency data from the dot probe task above is used to make an inference on attention bias indirectly. Using the eye tracker-generated data, I can explore attention one

Table 4.3: Models 1.0–1.2, Attentional bias (dot probe response latency)

	m1.0	m1.1	m1.2
DV		*Response latency, standardised (z) scale*	
Type	Linear / MCMC	Linear / MCMC	Linear / MCMC
Stacked N	700	700	599
Subject N	28	28	28
Trial N	20	20	20
DIC	1380.91	1389.59	1178.17
	Beta [95% CI]	Beta [95% CI]	Beta [95% CI]
Intercept	-0.06[-0.38;0.25]	-0.04[-0.34;0.27]	-0.04[-0.36;0.30]
30%		-0.07[-0.30;0.14]	-0.06[-0.27;0.18]
50%		-0.10[-0.37;0.13]	-0.08[-0.33;0.15]
60%		-0.05[-0.19;0.13]	-0.04[-0.20;0.14]
Contrast pic.		0.02[-0.09;0.15]	0.06[-0.08;0.22]
30% x Cntr		0.19[-0.08;0.49]	0.15[-0.13;0.44]
50% x Cntr		0.03[-0.30;0.37]	-0.03[-0.36;0.33]
60% x Cntr		-0.06[-0.30;0.18]	-0.09[-0.35;0.15]
	Gamma [95% CI]	Gamma [95% CI]	Gamma [95% CI]
Subject var (Random)	0.61[0.33;1.00]	0.60[0.30;0.94]	0.61[0.34;1.00]
trialnum	0.06[0.02;0.11]	0.06[0.02;0.11]	0.05[0.01;0.09]
Residual var	0.39[0.36;0.44]	0.40[0.35;0.44]	0.39[0.34;0.43]

step more directly, by looking at selected locations of visual attention, and quantify engagement with these locations by counting up the number and duration of total fixations—gaze kept on/within single location—and the number and duration of saccades—rapid eye movements—crossing them.

Inference from fixation duration measures is the inverse of inference from dot-probe response latency: difficult, cognitively complex, and/or attention grabbing stimuli should result in *greater length* of a single fixation [150]. It also makes sense to expect that fixations are more *frequent* on more engaging pictures while subjects explore more details within these images. I also analyse saccadic movement data, but this project is more exploratory. Saccades are involuntary movements between points of fixations, while information is typically acquired during fixations. While long and frequent saccadic movements may take time away from information processing, greater saccade frequency may also imply that subjects were exploring more details in the imagery.

Areas of Interests (AoIs) are fields of picture stimuli where fixation and saccade activity is measured, discarding gaze data on alternative fields. Following the logic of the dot-probe task, the AoIs are selected as to correspond to each picture across the picture pairs presented to participants through 20 trials. Using the log of the experimental trials[13], I matched these AoIs to the actual picture files presented and thus obtained, for each picture pair, an AoI featuring less women (or no women at all), an AoI featuring more women, and information on percentage of numerical balance on the pictures.

Based on the role model hypothesis on attention, the general expectation is that Role model (more women) AoIs get more attention than Contrast (less women) AoIs. Learning from the provisional dot-probe results earlier, however, it is further expected that Role

model AoIs get the attention surplus when featuring a single or token woman surrounded by larger group of men—a proportion of about 20%—but receive less attention if women do not 'stand out' in contrast to men (30% and beyond).

The unit of analysis is the image-area of interest, resulting in an $N = 2$ (images in a single pair) $\cdot$ 20 (trials, the number of pairs presented) $\cdot$ 18 (participants without missing data) $= 720$. Recording is entirely missing for two participants due to software failure (no video recorded during trial). There were tracking quality issues for three participants, I excluded those from further analysis. Another five participants did not follow the instructions to observe the pictures but, given their satisfactory dot probe task performance (two subjects made no errors at all, and three subjects made one error each across 20 trials), they may have adopted another strategy to detect the probes without. While it may seem like a substantial loss of data, due to the intensive nature of data collection with an eye tracking , I am still working with plenty of observations.

Though conceptually separate, data on the number of fixations, duration of fixations, number of saccades and duration of saccades overlap empirically, at least in terms of their distribution across treatment groups, as demonstrated in Table 4.2. The descriptive statistics report comparable movements to response latency, confirming basic validity of the dot probe task in measuring attentional bias, with a few exceptions.

First, gaze data detects somewhat greater range of treatment averages, while group means in response latency went from about 370 to 420 milliseconds, gaze duration covers group differences of up to 70 milliseconds and saccade duration of 750 milliseconds. Latter range is exaggerated as often saccades were not picked up on an AoI at all. Second, gaze data is simultaneously recorded on both pictures in pair, offering perhaps better within-subject controls, while a dot-probe response is recorded either after the role model picture, or after the Contrast picture. Third, of more substantive interest, there may be more validity behind a token effect as the attention gap is not only on the 20% pictures but also on the 25% picture, previously an outlier. Finally, looking at women majority pairings, the attention gap led by Contrast pictures, as observed in the dot-probe data, seems less exaggerated or disappears entirely suggesting that pictures where men became novelty may have received less attention than previously found.

Through exploratory data analysis, a pattern emerged in which probe location, that is, where the dot after picture stimuli was presented, is reliably and strongly associated with, above all, gaze duration ($t(719.01) = -32.39, p < .001$). This would mean that participants predicted where the probe was going to appear (by spending about 30 milliseconds more time on those pictures), which is an unlikely explanation given randomised trials. The log of the experimental trials confirms that randomisation was successful. Thus the only explanation left is that, though care was taken to extract gaze data for times only when pictures were on screen, the software picked up attention directed at the probes immediately the pictures disappeared for about 30 milliseconds. I am returning to this point in the section conclusions.

I fit hierarchical models following the same logic as explained above under Section

4.1.1. Duration of fixation and duration of saccades are scaled (mean centred, divided by standard deviation) to allow direct comparability to response latency results. I did not make changes in the hierarchical structure, nesting is within trials and individuals, and there will be no subsequent exclusion of the 25% picture AoI from the sample.

Fixation and saccade duration data models are linear regressions at the core fit with MCMC, comparable directly to the dot-probe model,

$$Y_{duration} = \beta_0 + u_{0t}^{trial} + u_{0s}^{subject} + \beta_1 X_t^{pic.\ type} \cdot \beta_2 X_t^{\%\ balance} + \beta_3 X_t^{probe\ loc.} + e \qquad (4.3)$$

with a familiar host of predictors comprising picture type (Contrast versus Role model), numerical balance in the Role model picture, but also new piece of information being lagged probe location as discussed above. *Count* data of fixations and saccades, on the other hand, follows a Poisson rather than standard normal distribution, which is accounted for in these models by using the appropriate link function:

$$\ln(Y)_{count} = \beta_0 + u_{0t}^{trial} + u_{0s}^{subject} + \beta_1 X_t^{pic.\ type} \cdot \beta_2 X_t^{\%\ balance} + \beta_3 X_t^{probe\ loc.} \qquad (4.4)$$

where the same linear combination of predictors designate the natural logarithm of the dependent variable (count of fixations). These effects sizes are reported in the appropriate model tables, while the prediction plot will be show values transformed back to on the original count scale.

Predicted duration of fixations
95% distribution of posterior predictions

Proportion of women in picture stimuli
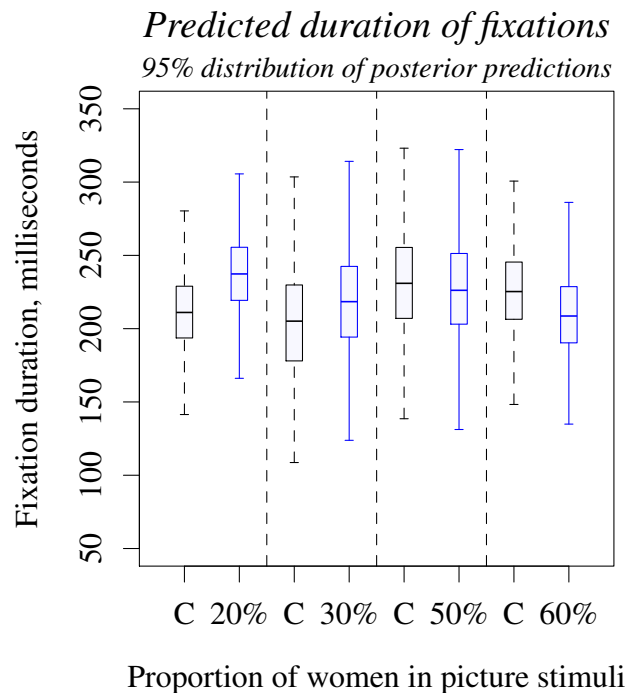
Note: C stands for 'Contrast image' in pair featuring less or no women. Longer fixations may imply attentional bias.

Figure 4.2: Attentional bias across picture contrasts, Model 2.2

My results reveals a pattern very similar if more moderate compared to the dot-probe results. Evidence behind the token effect weakens, likely due to a decrease in subject sample size. In Model 2.2, Contrast parameter (that is, the attention gap in contrast to role model pictures with a token or 20% presence only) remains negative, as expected, in a smaller 85% highest density interval: though still non-negligible probability. The treatment effect is identical to that in the dot-probe task before excluding the 25% picture (-.14 and .13, respectively, note that direction of inference should be opposite from response latency and fixation duration). The gap is still very small, if existing at all, between contrast pictures and 30–50% role model pictures. Finally, though the descriptives suggested no separation of contrast and role model pictures at women majority pairs, the model-based answer is similar to the dot-probe result in that attention is re-directed to men-in-minority pictures. Figure 4.2 shows these predicted probabilities.

Looking at variance components, two important results emerge. First, the uninformative model shows about 75% variability in fixation duration is unexplained by simply looking at 'blank' between-subject variation (which itself remains around 25% throughout model expansion) suggesting attentional engagement patterns may not be entirely up to individuals but it has to be a more elaborate mechanism in terms of variability in stimuli. Second, including gender balance predictors, however uncertain estimates in terms of posterior density, do decrease this unexplained variance about 7%, showing better explanatory power than lagged probe location, accounting for a mere 1%.

Fixation count results largely confirm patterns found in results about fixation duration. The magnitude of the token effect is comparable to that measured through fixation duration and response latency previously, the single standing Contrast coefficient remaining around -.17 or a fifth of the dependent variable's standard deviation. This parameter now has higher posterior density and remains negative within 95% credible intervals. This model however finds very weak evidence that these patterns may reverse when men become majority.

Figure 4.3 shows the posterior predictions transformed back on the original scale. In terms of fixation counts, the role model effect under women's token presence, though detected by the model, fails to account for one extra fixation, which may explain another piece of detail observed. I will discuss this issue again in the Section Conclusions.

Saccadic data models show weak evidence that Contrast pictures were implicated in fewer and shorter saccades. Saccadic duration effect is the strongest of the two. Saccades being involuntary movements between points of fixations, this duration effect may mean that quick saccades connected two fixation points *within the same pictures*—previous model confirmed that about every one in five token pictures had an extra fixation. The saccadic count effect, though positive, has very large posterior variance. This, too, is consistent with the fixation count model which found low predicted probability of fixation counts beyond 2, which would be necessary to detect multiple saccades between them. Across the two models, the remaining parameter estimates are inconsistent.
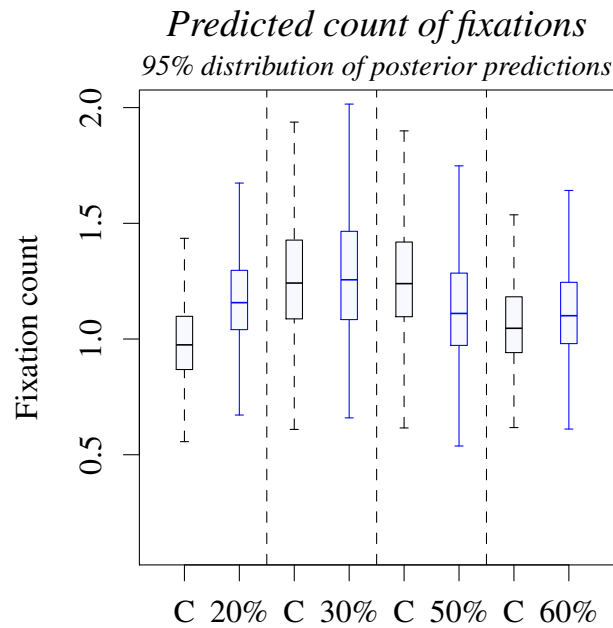
Figure 4.3: Attentional bias across picture contrasts, Model 3.2

4.1.3 Section conclusions

Using a strategy that seeks to get at a singular underlying concept through a number of alternative measures, namely attentional bias to political role models, I observe very consistent effects. When contrasted with men only groups, a token woman's novelty draws women participants' attention, consistently shown through performance on a cognitive task (the dot-probe task), as well as gaze measures. Role model pictures featuring the token woman were 'attended' areas of the picture contrasts and were followed by a fifth standard deviation quicker response time, about 33 milliseconds. These pictures also were implicated in a fifth standard deviation longer fixations, remaining above 30 milliseconds through all models. About every fifth role model picture may have received an extra fixation, suggesting that more information was taken in from token pictures. Despite magnitude differences, however, posterior probabilities vary, weakest treatment effects remaining non-zero within 85%, strongest within 95% credible intervals.

Though exploratory, I also analyse pairings with different numerical balance contrasts, finding that although role model pictures continue to get attention surplus, this may diminish by parity when women no longer stand out as minority. Rather, as men become minority in two picture pairings, participants redirected attention at them. My further exploratory results on saccadic data is consistent with the idea that participants re-focused on role model pictures, as my model found evidence that quicker saccades happened on role model pictures. As the probability of re-fixation was estimated to be very low in the previous models thus a third, fourth etc. fixations deemed very unlikely, saccade count model found no evidence of further saccadic activity on role model pictures.

Table 4.4: Models 2.0–2.2, Attentional bias (fixation duration)

	m2.0	m2.1	m2.2
DV	*Duration of fixation, standardised (z) scale*		
Type	Linear / MCMC	Linear / MCMC	Linear / MCMC
Stacked N	720	720	720
Subject N	18	18	18
Trial N	20	20	20
DIC	1869.32	1880.01	1877.91
	Beta [95% CI]	Beta [95% CI]	Beta [95% CI]
Intercept	0.00[-0.24;0.27]	0.09[-0.18;0.38]	0.02[-0.27;0.27]
30%		-0.12[-0.42;0.19]	-0.10[-0.39;0.21]
50%		-0.05[-0.34;0.24]	-0.06[-0.36;0.22]
60%		-0.15[-0.36;0.05]	-0.14[-0.36;0.06]
Contrast pic.		-0.14[-0.32;0.03]	-0.13[-0.30;0.06]
Probe lag			0.13[0.01;0.27]
30% x Cntr		0.10[-0.37;0.52]	0.06[-0.43;0.48]
50% x Cntr		0.14[-0.22;0.63]	0.15[-0.33;0.52]
60% x Cntr		0.23[-0.04;0.53]	0.21[-0.12;0.49]
	Gamma [95% CI]	Gamma [95% CI]	Gamma [95% CI]
Subject var (Random) var	0.24[0.11;0.50]	0.25[0.12;0.53]	0.25[0.12;0.53]
Trial var	0.00[0.00;0.00]	0.00[0.00;0.00]	0.00[0.00;0.00]
Residual var	0.77[0.69;0.85]	0.70[0.59;0.81]	0.69[0.59;0.85]

Table 4.5: Models 3.0–3.2, Attentional bias (fixation count)

	m3.0	m3.1	m3.2
DV	*Fixation count*		
Type	Poisson / MCMC	Poisson / MCMC	Poisson / MCMC
Stacked N	720	720	720
Subject N	18	18	18
Trial N	20	20	20
DIC	1917.95	1903.41	1913.54
	Beta [95% CI]	Beta [95% CI]	Beta [95% CI]
Intercept	0.20[-0.07;0.49]	0.18[-0.18;0.53]	0.17[-0.17;0.48]
30%		0.06[-0.20;0.33]	0.01[-0.30;0.30]
50%		-0.04[-0.22;0.16]	0.01[-0.24;0.25]
60%		-0.07[-0.23;0.07]	-0.13[-0.35;0.05]
Contrast pic.		-0.14[-0.32;0.00]	-0.13[-0.29;0.06]
Probe lag			0.06[-0.04;0.23]
30% x Cntr		0.08[-0.23;0.52]	0.20[-0.19;0.59]
50% x Cntr		0.32[0.04;0.51]	0.20[-0.23;0.68]
60% x Cntr		0.00[-0.26;0.23]	0.17[-0.08;0.40]
	Gamma [95% CI]	Gamma [95% CI]	Gamma [95% CI]
Subject var (Random) var	0.35[0.15;0.63]	0.52[0.22;0.95]	0.46[0.17;0.84]
Trial var	0.00[0.00;0.00]	0.00[0.00;0.00]	0.00[0.00;0.00]

Table 4.6: Models 4.0–4.2, Attentional bias (duration of saccades)

	m4.0	m4.1	m4.2
DV	*Duration of saccades, standardised (z) scale*		
Type	Linear / MCMC	Linear / MCMC	Linear / MCMC
Stacked N	720	720	720
Subject N	18	18	18
Trial N	20	20	20
DIC	1959.46	1969.07	1970.5
	Beta [95% CI]	Beta [95% CI]	Beta [95% CI]
Intercept	0.00[-0.18;0.19]	0.05[-0.18;0.29]	0.04[-0.17;0.28]
30%		-0.02[-0.35;0.29]	-0.02[-0.38;0.30]
50%		0.01[-0.33;0.30]	0.00[-0.33;0.29]
60%		0.04[-0.17;0.27]	0.03[-0.17;0.26]
Contrast pic.		-0.12[-0.32;0.07]	-0.12[-0.30;0.07]
Probe lag			0.01[-0.11;0.15]
30% x Cntr		0.15[-0.35;0.60]	0.14[-0.33;0.63]
50% x Cntr		0.16[-0.29;0.58]	0.17[-0.26;0.59]
60% x Cntr		-0.10[-0.43;0.23]	-0.09[-0.41;0.19]
	Gamma [95% CI]	Gamma [95% CI]	Gamma [95% CI]
Subject var (Random)	0.16[0.05;0.29]	0.16[0.06;0.30]	0.15[0.05;0.27]
Trial var (Random)	0.00[0.00;0.00]	0.00[0.00;0.00]	0.00[0.00;0.00]
Residual var	0.87[0.78;0.96]	0.87[0.77;0.96]	0.87[0.79;0.97]

Table 4.7: Models 5.0–5.2, Attentional bias (count of saccades)

	m5.0	m5.1	m5.2
DV	*Count of saccades*		
Type	Poisson / MCMC	Poisson / MCMC	Poisson / MCMC
Stacked N	720	720	720
Subject N	18	18	18
Trial N	20	20	20
DIC	1420.79	1429.62	1411.51
	Beta [95% CI]	Beta [95% CI]	Beta [95% CI]
Intercept	-0.65[-0.99;-0.39]	-0.61[-0.97;-0.25]	-0.73[-1.07;-0.37]
30%		0.02[-0.56;0.44]	0.01[-0.16;0.24]
50%		-0.01[-0.39;0.39]	0.08[-0.11;0.38]
60%		-0.09[-0.39;0.24]	-0.12[-0.31;0.10]
Contrast pic.		-0.23[-0.46;0.02]	-0.17[-0.28;-0.04]
Probe lag			0.19[0.09;0.27]
30% x Cntr		0.43[-0.11;1.14]	0.30[0.00;0.59]
50% x Cntr		0.40[-0.14;1.03]	0.22[-0.30;0.48]
60% x Cntr		0.07[-0.46;0.49]	0.03[-0.32;0.27]
	Gamma [95% CI]	Gamma [95% CI]	Gamma [95% CI]
Subject var (Random)	0.39[0.13;0.72]	0.44[0.17;0.83]	0.54[0.20;1.02]
Trial var (Random)	0.00[0.00;0.00]	0.00[0.00;0.00]	0.00[0.00;0.00]
Residual var	0.04[0.00;0.12]	0.05[0.00;0.13]	0.00[0.00;0.01]

4.2 Political engagement

This section reviews evidence of role model effects more directly on psychological engagement with politics, relying on two data sources. First, I collected original data in a survey experiment where respondents were randomly assigned to view images of political groups depicting varying numerical balance of men and women. This project comprises two data collection rounds, an 'original', $N = 329$ study fielded May 2015, and a replication round in June 2017, where $N = 277$ pariticpants[14] received the original treatment unchanged.

Second, I use observational data from the British Election Study projects, namely, an $N = 15,658$ internet panel [127] and an $N = 1,935$ face to face panel study before and after the election campaign of 2010 [126], and an $N = 31,323$ internet panel before and after the election campaign of 2015[128]. Most of the variables of interest are available in the large internet samples, while the smaller 2010, face-to-face panel study is used only to explore if the women's campaign impacted trust in politicians. I link the BES data, where respondents' parliamentary constituency identifier is available, with data provided by the Houses of Parliament's Digital Service [125] on the candidates running in the respective constituencies. While winning majority and candidate names are readily available for these election years, candidate gender had missing data for 2010. Wherever the same person, 2015 candidate gender is imputed on these places. Further missing data was filled in using crowd-sourced information site Democracy Club[151], the 'babynames' data set relying on historical US Social Security Number Application data to predict most probable gender given first name, and finally, hand coding residuals[15].

Both BES projects ran across many waves with a good number of respondents staying in the panels, the waves being numbered differently. The 2010 face-to-face panel only had a pre-election and post-election wave, did not implement a campaign wave between the two. To avoid confusion, for all observational study, Wave 1 will refer to a pre-election wave, in the case of the BES 2015, the the one closest to but still preceding the campaign; and Wave 3 will refer to the post-election wave, the one closest to the actual vote cast (typically asked a day after). Wave 2 might be reserved for the internet panel 'campaign waves' in a possible extension of this study, but not currently used due to its diffuse timing and difficult comparison between the two election years.

The basic analytical strategy was to take measures of a wide range of indicators of psychological engagement with politics: political interest, political knowledge, and political efficacy; and assess differences under the following numerical balance setups in the picture or campaign contexts: 0% or *'men only'* politics, 20-25% or *token*, near 30% *critical mass*, and a 50% thus *parity* of men and women. These setups are readily available as they correspond to the treatment groups in the survey experiment, while they are calculated for the observational study based on the number of women standing as candidates over the total number of candidates. I found that 272 races featured one woman, or less than a third, in 2015 and 282 in 2010. 122 (2010) and 169 (2015) races more women

but less than 50%, 35 (2010) and 32 (2015) races an equal number of women and men.[16] An extra group denotes *women majority* presence in the campaign context (66 races) but the comparable category is missing from the pool of pictures. Table 4.8 below shows the number of races by women's presence.

Table 4.8: *N* of constituencies by women's campaign presence, GE 2010 & 2015

No. women *Treatment ↓*	0	1	2	3	4	5	6
2010							
M.only	146						
Token		237	64	1			
C.mass			76	43	3		
Parity			7	22	4	2	
W.maj				8	12	4	1
2015							
M.only	101						
Token		223	62	2			
C.mass			116	53		1	
Parity			2	24	6		
W.maj				2	2	2	

M.only = no women ran. Token = 1 woman race but always less than 30% of all candidates. C.mass races surpass 30% but equal to less than fifty-fifty, Parity. W.maj = women majority races.

Participants are relatively evenly distributed across treatment groups in the survey experiment due to random assignment. By contrast, respondents of the 2015 election study are highly unevenly distributed across race type, the vast majority of them being respondents under a token woman presence, with 'men only' and critical mass races coming up second most populous, but already with half as many respondents, while very few respondents observed parity races (Table 4.9 at Chapter appendix shows distribution across treatment groups for both datasets). Evidently latter distribution follows the actual distribution of gender balance across the 614 parliamentary constituencies sampled.

Table 4.9: *N* of respondents, by data source & by treatment—gender combination

Treatm.	Gender	BES 2010	BES 2015	Exp Round 1	Exp Round 2
M.only	Men	1970	2385	41	38
	Women	1991	2456	51	41
Token	Men	4257	7109	41	36
	Women	4342	7440	39	31
C.mass	Men	1615	4166	40	35
	Women	1703	4278	47	28
Parity	Men	499	844	38	31
	Women	480	894	32	36
W.maj.	Men	347	815		
	Women	389	936		

The dependent variables are change in *election interest* from baseline, Wave 1 to Wave 3, in BES 2010 and 2015; *political interest*, as reported by survey participants, survey experiment; change in self-assessed or *subjective political knowledge*, Wave 1 to Wave 3,

in BES 2015; self-assessed or *subjective political knowledge*, survey experiment; change in performance on *political knowledge test* Wave 1 to Wave 3, in BES 2015; performance on *political knowledge test*, survey experiment; *confidence* in political knowledge test answers, survey experiment; change in *political efficacy*, Wave 1 to Wave 3, in BES 2010 and 2015; *political efficacy*, survey experiment; change in *trust in MPs*, Wave 1 to Wave 3, in BES 2010 and 2015; and *Trust in MPs*, survey experiment.

Although the empirical models will vary based on measurement levels across the dependent variables, I am aiming to keep the structure of predictors fixed. The most important independent variable is treatment's (featuring the 4-5 numerical balance setups) interaction term with respondent gender, resulting in 8 to 10 point predictions. Furthermore, in models using survey experiment data, I fit hierarchical models to include a random term for the crowd-sourcing platform through which respondents were recruited, to account for any variation due to average respondent characteristics between these platforms. Thus for any linear dependent variable of psychological engagement with politics, Y_{PEP}, my models take a generic form,

$$Y_{\text{PEP}} = \beta_0 + u_{0p}^{platf.} + u_{0s}^{subject} + \beta_1 X_s^{resp.\ gender} \cdot \beta_2 X_s^{\%\ treatment} + e \qquad (4.5)$$

where $u_{0s}^{subject}$ is optional random effect if the dependent variable is multi-response variable nested within respondents, as explained earler at the introduction and at equation 4.1.

The random term for crowd-sourcing platform is also a component that allows me to disregard further covariates from the survey experiment models to 'save' on test power, needed in a small N experiment. There are very minimal 'within-treatment' differences in the distribution of key demographics such as education and age due to the element of random assignment to treatment groups, and differences are mostly on gender (which variable is present in all equations).

Latent variable models include an additional random term. When looking at multi-item dependent variables, such as political efficacy measured across four items, I pool these resopnses together such as the unit of analysis becomes the item response, stacked within individuals thus an item-level or stacked sample size equals to $N_{\text{item}} \cdot N_s$ where s stands for subjects.

$$Y_{\text{PEP}} = \beta_0 + u_{0p}^{platf.} + u_{0s}^{subject} + u_{0i}^{item} + \beta_1 X_s^{resp.\ gender} \cdot \beta_2 X_s^{\%\ treatment} + e \qquad (4.6)$$

In this equation, Y_{PEP} is latent construct, whereas u_{0i}^{item} gives variation across items within this latent construct.

In the election study data, 'treatment' varies on the constituency level but not on the respondent level, thus models with this data source will further include a random component for parliamentary constituency ID. Furthermore, since an element of randomisation is entirely missing in observational studies, I include further covariates in these models

to exclude effects reflecting on alternative, individual-level explanations. Keeping them constant, I choose a minimal, most stable set of predictors of political interest[9],[7], including age and 6 formal education levels ranging from 'None' to 'Postgraduate'. Furthermore, to account for constituency-level explanations other than the density of female politicians, I include winning majority, to control for how close the race might have been. The assumption is that races where candidates win by large majorities may be considered safe or uninteresting seats, resulting in less intense campaigning.

For any linear dependent variable that is a *change* of psychological engagement with politics during (women's) campaign, ΔY_{PEP}, my models take a generic form,

$$
\begin{aligned}
\Delta Y_{\text{PEP}} =\; & \beta_0 + u_{0c}^{constituency} + u_{0y}^{year} + u_{0s}^{subject} + \\
& \beta_1 X_s^{resp.\ gender} \cdot \beta_2 X_s^{\%\ treatment} + \\
& \beta_3 X_s^{age} + \beta_4 X_s^{education} + \beta_5 X_s^{majority} + e
\end{aligned}
\tag{4.7}
$$

where $u_{0s}^{subject}$ is optional random effect if the dependent variable is multi-response variable nested within respondents, such political efficacy.

The following sections review the results of these models, grouped by dependent variables. All hierarchical models are fit with MCMC using an uninformative or flat prior[17], and are built up from a 'null' model made up of a grand mean and variance components, expanded to include gender, treatment, and finally covariates (if observational study). Further details on the hierarchical structure such as for latent models (political knowledge and efficacy) are provided under each model table. I concentrate on major findings and plot marginal effect sizes, and make a note of the substantively interesting variance components to infer explanatory power or 'model fit'. Details on model convergence and a plot of each parameter's posterior distribution are provided in the Chapter appendix.

4.2.1 Political interest

In both the survey experiment and the BES, political interest is assessed on Likert-type scales. While in the survey experiment, I asked participants about 'general' political interest, the BES panel waves both in 2010 and in 2015 asked participants about election interest. For exploration purposes, Tables 4.10–4.11 show the simple group means per data set. Note that only very small changes are expected: previous studies have shown that political interest is fairly stable over the life course [152], therefore I am looking for only marginal campaign effects at best. The observed means through the BES panel waves show that in both election years, survey respondents became more interested by the end of the election campaign period, and that the magnitude of this change is always larger for women. $M\Delta$ shows the averaged 'within-individual' change suggesting very small differential across 'treatment' groups; the magnitude of gains nevertheless seeming largest in constituencies with a token woman campaigning in 2015, and a critical Mass campaigning in 2010. These averages, however, may hide further heterogeneity as campaign interest is expected to be conditional foremost on individual characteristics, such

as age and education, but also contextual characteristics, such as the closeness of the race. Dispersion is near-identical across treatment groups and election years, but somewhat larger for women than for men, standard deviations ranging between the extremes of $SD = 0.72$ and $SD = 0.89$.

Table 4.10: Descriptive statistics: Campaign interest; BES 2010

Gender	Treatm.	M Wave 1	M Wave 3	$M\Delta$
Men	M.only	3.41	3.55	0.13
	Token	3.39	3.53	0.11
	C.mass	3.38	3.52	0.12
	Parity	3.43	3.58	0.12
	W.maj.	3.43	3.58	0.09
Women	M.only	3.14	3.40	0.22
	Token	3.14	3.40	0.23
	C.mass	3.11	3.39	0.26
	Parity	3.21	3.45	0.24
	W.maj.	3.23	3.52	0.24

Table 4.11: Descriptive statistics: Campaign interest BES 2015

Gender	Treatm.	M Wave 1	M Wave 3	$M\Delta$
Men	M.only	3.51	3.56	0.05
	Token	3.51	3.55	0.03
	C.mass	3.50	3.55	0.04
	Parity	3.53	3.56	0.03
	W.maj.	3.52	3.55	0.03
Women	M.only	3.26	3.40	0.12
	Token	3.25	3.40	0.15
	C.mass	3.25	3.39	0.13
	Parity	3.28	3.43	0.13
	W.maj.	3.26	3.41	0.13

In the original and replication rounds of the survey experiment, I tested different scales (4 and 5 points, respectively) for more potential comparison with different observational studies, but the last columns of Table 4.12, 'standardised (z)' makes them comparable, expressing differences from the overall mean (set 0) in standard deviation units. In the first round of data collection, survey experiment results suggest that women's political interest is highest in the token condition, a significant increase from the baseline, Men only condition, but as the proportion of women in the pictures increased, mean reported political interest decreased somewhat (although not again hitting baseline Men only levels). The replication round, however, showed different patterns: women's political interest decreased in all treatment conditions, except parity, suggesting large within-treatment variances.

In the election interest models, Models 6.0–6.2, the random component for Parliamentary constituency bears substantive interest as it is an estimate of the variance found across constituencies, at which level the number of women campaigning varies. A large variance component may thus facilitate large substantive 'treatment' effects, however, the null model shows that constituency-level variation between respondents' reported change in

Table 4.12: Descriptive statistics: Political interest; Survey experiment

Gender	Treatm.	*M* Round 1	*M* Round 2	*M* Round 1 standard (z)	*M* Round 2 standard (z)
Men	M.only	3.02	3.79	0.21	0.06
	Token	2.95	4.06	0.11	0.33
	C.mass	2.95	3.91	0.11	0.19
	Parity	2.79	3.74	-0.09	0.01
Women	M.only	2.67	3.71	-0.24	-0.02
	Token	2.97	3.61	0.14	-0.12
	C.mass	2.81	3.07	-0.07	-0.66
	Parity	2.75	3.78	-0.14	0.05

campaign interest is essentially zero. In the first expanded model, 6.1, where respondent gender is included as a predictor, a little space may have opened up for constituency-level variation but remained almost negligible (0.001).

Nevertheless, Model 6.2 including the treatment interaction and the full host of controls, does find evidence that, everything else held equal,

1. Women's, rather than men's, election interest increased through both campaign periods, see Beta 'Gender: Woman' in Table 4.13, and the 'W' marginal effect in Figure 4.4;

2. Men were not sensitive to the gender composition of the campaigns;

3. The token woman campaign presence marginally increased women's election interest, on top of the baseline gender effect, if only by less than a hundredth standard deviation, see Beta 'Woman x 20%' or 'Wx20' in Figure 4.4;

4. Beyond token, there are no more marginal increases in election interest.

Gender, along with the proportion of women in campaigns, explains about one per cent variability of election interest, see decreasing Residual variance component. The inclusion of covariates has somewhat larger explanatory power, with about 5% variability now explained by adding age, education, and candidates' winning majority. Interestingly, winning parties' vote share is found to have an impact on political interest in the unexpected direction. Though a tiny effect size, results suggest that a growing winning percentage (an indication of a 'safe seat') translates to an increased political interest. The top panel of Figure 4.4 shows these effects expressed in terms of standard deviation units of the dependent variable. The full results are in Table 4.13.

Models 7.0–7.2 on political interest based on the survey experiment data find generally no treatment effects. A token boost comparable to the BES model, though observed in the first round of data collection, is now counter-balanced by negative movements in the replication study. Model expansion to respondent gender and treatment group does little to reduce individual-level, unexplained variance (Residual variance component), pointing to a very weakly informative model. The bottom panel of Figure 4.4 shows key marginal

Figure 4.4: Treatment x gender marginal effects, Models 6.2 & 7.2
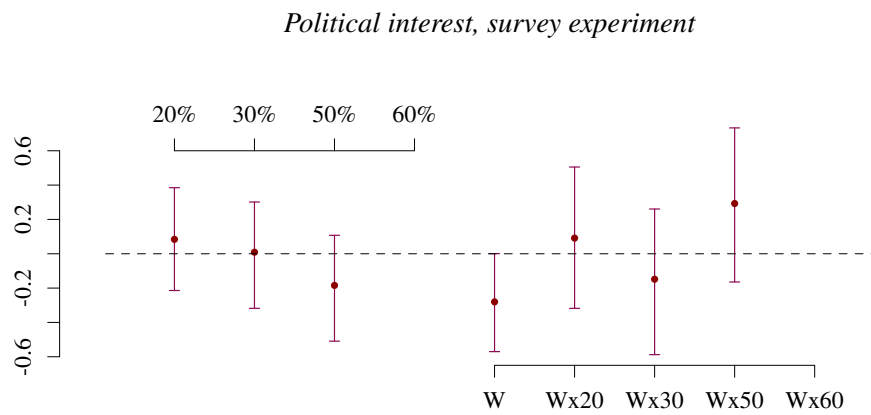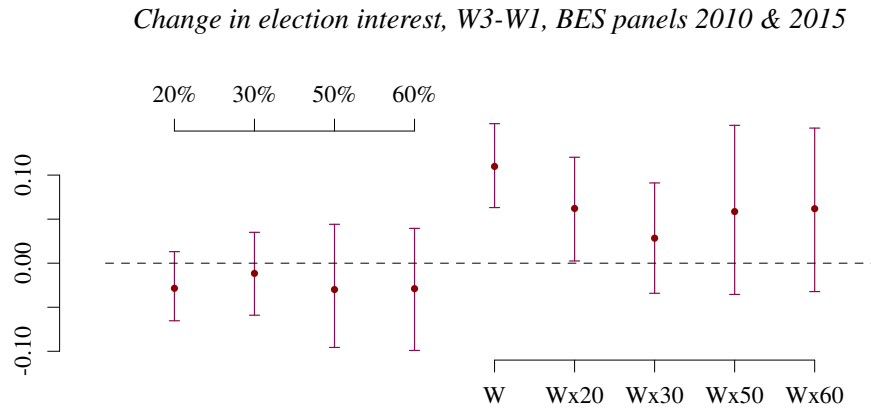
Change in election interest, W3-W1, BES panels 2010 & 2015

Political interest, survey experiment

Key: β coefficients (posterior means) and 95% CI corresponding to the effect of women's presence in campaigns (BES) or picture stimuli (survey experiment) at the discrete proportions of near 20% or token, near 30% or critical mass, 50% or parity, and women's majority near 60%. W stands for women respondents.

effects, with the full results in Table 4.13. Model convergence trace plots are in Section 4.4 Chapter appendix, introduced by general remarks.

Table 4.13: Models 6.0–6.2, Change in election interest, BES

	m6.0	m6.1	m6.2
DV	*Election interest Wave 3–Wave1, standardised (z) scale*		
Type	Linear / MCMC	Linear / MCMC	Linear / MCMC
N	48916	48916	36737
Constituency N	632	632	632
Election years	2	2	2
DIC	111070.4	110805.53	102389.28
	Beta [95% CI]	Beta [95% CI]	Beta [95% CI]
Intercept	0.02[-0.72;0.90]	0.09[-0.92;1.39]	-0.02[-0.62;0.71]
Gender: Woman		0.13[0.08;0.18]	0.11[0.07;0.16]
Treatment: 20%		-0.02[-0.06;0.02]	-0.03[-0.06;0.01]
Treatment: 30%		0.00[-0.05;0.04]	-0.01[-0.06;0.03]
Treatment: 50%		-0.02[-0.09;0.05]	-0.03[-0.10;0.04]
Treatment: 60%		-0.03[-0.10;0.04]	-0.03[-0.09;0.04]
Woman x 20%		0.05[0.00;0.11]	0.06[0.01;0.12]
Woman x 30%		0.02[-0.04;0.08]	0.03[-0.03;0.09]
Woman x 50%		0.05[-0.05;0.14]	0.06[-0.04;0.15]
Woman x 60%		0.05[-0.04;0.15]	0.06[-0.03;0.15]
Resp.: Age			-0.09[-0.10;-0.08]
Resp.: Educ.			0.01[0.00;0.02]
Const.: Marginality			0.01[0.00;0.02]
	Gamma [95% CI]	Gamma [95% CI]	Gamma [95% CI]
Constit. var (Random)	0.00[0.00;0.00]	0.00[0.00;0.00]	0.00[0.00;0.00]
Election year var (Random)	1.67[0.00;2.16]	7.10[0.00;4.42]	11.23[0.00;1.00]
Residual var	1.00[0.98;1.01]	0.99[0.97;1.00]	0.95[0.94;0.96]

Table 4.14: Models 7.0–7.2, Political interest, Survey experiment

	m7.0	m7.1	m7.2
DV	*Political interest, standardised (z) scale*		
Type	Linear / MCMC	Linear / MCMC	Linear / MCMC
N	605	605	605
Platforms	2	2	2
Coll. round	2	2	2
DIC	1718.91	1712.33	1717.49
	Beta [95% CI]	Beta [95% CI]	Beta [95% CI]
Intercept	-0.10[-0.08;0.09]	0.12[0.01;0.24]	0.14[-0.10;0.34]
Gender: Woman		-0.24[-0.39;-0.08]	-0.29[-0.57;0.00]
Treatment: 20%			0.08[-0.20;0.40]
Treatment: 30%			0.01[-0.30;0.32]
Treatment: 50%			-0.19[-0.51;0.10]
Woman x 20%			0.10[-0.32;0.51]
Woman x 30%			-0.16[-0.59;0.26]
Woman x 50%			0.29[-0.17;0.72]
	Gamma [95% CI]	Gamma [95% CI]	Gamma [95% CI]
Coll.round var (Random)	0.00[0.00;0.00]	0.00[0.00;0.00]	0.01[0.00;0.00]
Crowds-pl. var (Random)	4.81[0.00;0.01]	0.02[0.00;0.00]	0.00[0.00;0.00]
Residual var	1.00[0.89;1.11]	0.99[0.88;1.10]	0.99[0.88;1.10]

4.2.2 Political knowledge

Political knowledge is introduced as relevant aspect of political engagement on at least three accounts. Firstly, greater stores of political knowledge, especially about current affairs, may be an indication that citizens follow politics. Secondly, the politically aware is more likely to be politically engaged in terms of participation [153],[154]. Thirdly, a subjective, *confident* assessment of one's political skills is part of her more holistic judgement self-efficacy [48],[49], measures I discuss separately under political efficacy.

Objectively, political knowledge is typically measured through factual political knowledge quizzes, while subjectively, it is assessed on Likert scales as part of a longer battery of questions about political efficacy. The BES panel of 2010 implemented none of these options, while the BES 2015 has a number of waves experimenting with different options. Participants were tested for their knowledge of current affairs in almost all campaign waves, but only three questions from before the election campaign were re-asked in the post-election wave. These were job pairing tasks where respondents had to match the names of the President of Russia, the Chancellor of Germany, and the President of Syria.

Table 4.15 on test performance and learning shows large gender gaps. For each of the three test questions, I determined if respondents answered correctly in the pre- and the post-election wave, and categorised them depending on whether they learned about the leaders or not, kept existing knowledge, or 'forgot' (gave correct answer in Wave 1 but an incorrect answer in Wave 3). Taking the most typical value across three test questions, '*N* Learned' counts the number of respondents who gave a correct response in Wave 3 but not Wave 1. 'Learned odds' shows the odds of this over 'Not learning' (both waves incorrect or 'don't know' response), suggesting that men were up to 4.5 times more likely to learn through the campaign than women, the odds being highest in men only races. Women had near equal odds in men only and token races but less in Critical mass and parity races, at least as far as the test question medians are concerned. The model-based solution will look at common learning patterns about all three leaders, not relying on the mode across three questions.

Table 4.15: Descriptive statistics: Change in knowledge of foreign leaders by Wave 3, BES 2015

Gender	Treatm.	*N* Not learn	*N* Learned	*N* 'Forgot'	*N* No change	Learn odds
Men	M.only	16	144	0	82	9.00
	Token	53	377	8	219	7.11
	C.mass	34	269	4	131	7.91
	Parity	7	46	1	23	6.57
	W.maj.	10	58	0	26	5.80
Women	M.only	110	255	6	89	2.32
	Token	300	636	26	234	2.12
	C.mass	195	386	14	112	1.98
	Parity	43	74	3	22	1.72
	W.maj.	29	70	5	24	2.41

Through the original and the replication rounds of the survey experiment, I adminis-

tered tests comprising 12 and 10 questions, respectively. The questions in the first round test civic knowledge and familiarity with the constitution, adopted from Audit of Political Engagement Survey [155], whereas the questions for the replication round resemble more the BES-style leaders and names matching exercise, emphasizing more current affairs. Table 4.16 shows average scores, confirming the gender gap on the (harder) matching task but not confirming the gender gap on the (easier) civic knowledge. Women did not seem to have been impacted by the treatment. For list of test questions, see Chapter 3 Section 3.3 Chapter appendix.
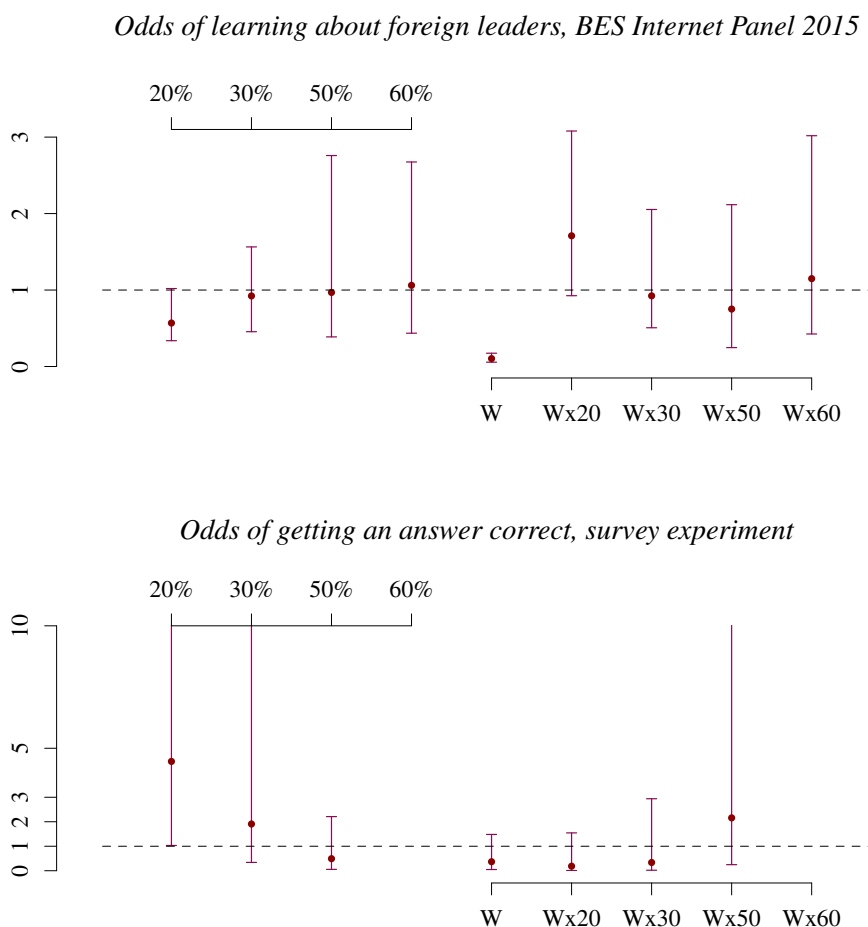
Table 4.16: Descriptive statistics: Political knowledge test scores; % correct

Gender	Treatm.	*M* score BES '15	*M* score Exp. Round 1	*M* score Exp. Round 2
Men	M.only	0.91	0.70	0.69
	Token	0.91	0.74	0.76
	C.mass	0.91	0.73	0.71
	Parity	0.91	0.72	0.65
	W.maj.	0.90		
Women	M.only	0.74	0.71	0.63
	Token	0.74	0.72	0.61
	C.mass	0.74	0.72	0.59
	Parity	0.75	0.73	0.64
	W.maj.	0.74		

Both exercises were paired with a rating task about confidence in the individual answers, analysed separately in the next section.

The BES panel models are Models 8.0–8.2 in Table 4.17 predicting the odds of learning about foreign leaders. The number of random components are extended to include test Question and Subject variance components: in these models, all answers are pooled together to be nested within the three test questions, and those within individuals. The model being predictive of odds, rather than a standardised-continuous dependent variable as before, variance components reflect how much variation in terms of (log) odds is due to nesting. Test question variance is very large[18], meaning that whether or not people learned about them is almost entirely up to their saliency or obscurity in the campaign. During the 2015 campaign, 38% of all panel respondents learned the name of the President of Syria, compared with 56% learning about the Chancellor of Germany, which may reflect major party priorities in campaign communication, at that time more focusing on the European Union. Between-individual variance is naturally large as learning may be more up to skills and motivation if embedded within similar information environments. Through model expansion, however, this 'subject' variance decreases, as we add information about education, age, and importantly, gender. Finally, in these models, between-constituency variance, on which level the proportion of women in campaigns is measured, is small but non-negligible. Constituency Gamma in Table 4.17 starts at 0.36, meaning that a variability of $e^{0.36} = 1.43$ in terms of odds may be up to any kind of constituency characteristics. For example, out of two constituencies with a population of 100 people, if 10 have learned about leaders in constituency A, 15 may be expected to learn in constituency B. Residual variance is not interpreted in probabilistic models.

Figure 4.5: Treatment x gender marginal effects, Models 8.2 & 9.2

Odds of learning about foreign leaders, BES Internet Panel 2015

Odds of getting an answer correct, survey experiment

Key: β coefficients (posterior means) and 95% CI corresponding to the effect of women's presence in campaigns (BES) or picture stimuli (survey experiment) at the discrete proportions of near 20% or token, near 30% or critical mass, 50% or parity, and women's majority near 60%. W stands for women respondents.

The results indicate that women are less likely to have learned about the three political leaders. In the full model specification, the odds of women's learning in constituencies with only men campaigning is $e^{-2.28} = 0.10$ or a tenth against men's learning. In terms of 'treatment' effects, somewhat over 90% of the posterior distribution points to a positive token effect on women's learning, the odds now being $e^{0.52} = 1.68$. Certainty about effects beyond token decreases in all setups. The top panel of Figure 4.5 shows the key marginal effects, while Table 4.17 shows the full results including control variables, confirming expectations.

Test score models 9.0–9.2 on the probability of a correct response using survey experiment data are in Table 4.18. 21 unique test answers— having implemented quizzes with 10 and 11-questions in the two data collection rounds, respectively—are submitted by 605 test subjects. Question-level variance is large, likely due to variation in question difficulties. For example, using the confidence ratings analysed in the next section, I learned that 84.41% of all respondents knew the Leader of the Labour Party with maximum confidence, but only 8.09% picked the Prime Minister of Italy with maximum confidence—a difficult item given especially that I included two Italian names in the pool of options. There is also large subject level variation, skills (and maybe confidence) exhibiting a

variation of $e^{1.79} = 5.47$ in terms of odds: this might be the space 'open' to be explained in terms of gender as well as exposure to gender balance in picture stimuli.

Gender does matter. Women fare less well on the tests, the odds estimated to be $e^{-0.36} = 0.70$ or a third less likely than men to get a correct answer. Model expansion including the effect of treatment, however, finds no reliable treatment effects on women. Men may have experienced a token boost in getting a correct answer, being 1.7 times more likely to get the correct answer if viewing pictures with a token woman than men looking at men only pictures, the effect remaining positive within 90% credible intervals. Men's odds may be gradually decreasing, but the effect sizes sit within wide CIs. The bottom panel of Figure 4.5 shows the relevant marginal effects.

As mentioned earlier, political knowledge may be assessed subjectively. Since conceptually very close to internal political efficacy, as argued at the beginning of this section and in Chapter 2 Section 2.1.1; as well as empirically part of most political efficacy batteries of survey items, such as the one asked in the BES, I will look at role model effects on a positive or self-confident assessment of one's political knowledge in the next section.

Table 4.17: Models 8.0–8.2, Log odds of learning about foreign leaders, BES 2015

	m8.0	m8.1	m8.2
DV	*Learning versus not learning about leaders by Wave 3*		
Type	Binomial / MCMC	Binomial / MCMC	Binomial / MCMC
Stacked N	8799	8799	8799
Question N	3	3	3
Subject N	3328	3328	3328
Constituency N	605	605	605
DIC	7076.47	6773.09	6572.14
	Beta [95% CI]	Beta [95% CI]	Beta [95% CI]
Intercept	1.74[-2.78;5.36]	2.84[-0.81;6.92]	2.98[-0.66;7.74]
Gender: Woman		-1.84[-2.35;-1.37]	-2.28[-2.87;-1.75]
Treatment: 20%		-0.19[-0.76;0.45]	-0.56[-1.08;0.02]
Treatment: 30%		-0.13[-0.86;0.45]	-0.11[-0.74;0.48]
Treatment: 50%		-0.49[-1.31;0.51]	-0.03[-0.99;0.96]
Treatment: 60%		-0.04[-0.76;0.67]	0.05[-0.86;0.94]
Woman x 20%		0.08[-0.61;0.64]	0.52[-0.11;1.07]
Woman x 30%		-0.10[-0.78;0.64]	-0.06[-0.71;0.67]
Woman x 50%		-0.10[-0.94;0.98]	-0.30[-1.33;0.77]
Woman x 60%		0.06[-0.74;1.02]	0.14[-0.89;1.07]
Resp.: Age			0.97[0.85;1.10]
Resp.: Educ.			0.53[0.43;0.64]
Const.: Marginality			-0.05[-0.15;0.06]
	Gamma [95% CI]	Gamma [95% CI]	Gamma [95% CI]
Question var (Random)	32.49[0.42;57.09]	14.85[0.29;47.65]	49.28[0.38;55.51]
Subject var (Random)	7.12[5.46;8.69]	5.36[4.40;6.07]	5.93[4.67;6.92]
Constit. var (Random)	0.36[0.09;0.70]	0.38[0.18;0.59]	0.29[0.02;0.52]
Residual var	0.06[0.01;0.10]	0.01[0.00;0.03]	0.02[0.01;0.05]

Table 4.18: Models 9.0–9.2, Political knowledge test scores, survey experiment

	m9.0	m9.1	m9.2
DV	*Probability of correct test answer*		
Type	Binomial / MCMC	Binomial / MCMC	Binomial / MCMC
Stacked N	6708	6708	6708
Test question N	22	22	22
Subject N	605	605	605
Coll. rounds	2	2	2
Platform N	2	2	2
DIC	6191.2	6205.5	6150.77
	Beta [95% CI]	Beta [95% CI]	Beta [95% CI]
Intercept	1.83[0.95;2.73]	1.84[1.06;2.58]	1.95[1.01;3.13]
Gender: Woman		-0.36[-0.63;-0.15]	-0.25[-0.80;0.22]
Treatment: 20%			0.53[-0.08;1.05]
Treatment: 30%			0.30[-0.19;0.91]
Treatment: 50%			-0.10[-0.63;0.48]
Woman x 20%			-0.54[-1.31;0.24]
Woman x 30%			-0.32[-1.17;0.37]
Woman x 50%			0.21[-0.63;1.00]
	Gamma [95% CI]	Gamma [95% CI]	Gamma [95% CI]
Question var (Random)	2.55[0.83;4.77]	2.08[0.82;3.60]	2.80[0.89;5.70]
Subject var (Random)	1.57[0.84;2.34]	1.26[0.93;1.66]	1.70[0.91;2.78]
Coll.round var (Random)	0.00[0.00;0.00]	0.01[0.00;0.00]	0.00[0.00;0.00]
Crowds-pl. var (Random)	0.03[0.00;0.02]	0.01[0.00;0.01]	0.62[0.00;0.05]
Residual var	2.50[0.23;5.01]	1.51[0.77;2.19]	2.96[0.45;6.14]

4.2.3 Political efficacy & related measures

Before analysing political efficacy more comprehensively, across *confidence about political knowledge* and *beliefs about influencing political outcomes* through participation, I will look at these components separately, subject to availability in the data sets. My dependent variables comprise, firstly, a single-item subjective assessment of political knowledge (BES 2015 and survey experiments); secondly, multi-item confidence measures developed for the survey experiment asking test subjects to indicate their level of confidence after each submitted answer; thirdly, a single item self-assessed political influence question asked in the 2010 BES, and finally, a multi-item political efficacy scale asked in the BES and mirrored in my survey experiment.

Subjective political knowledge

In the BES panel of 2015, the subjective self-assessment question was asked as part of a standard 'political efficacy' battery, where participants could express their agreement with the statement 'I understand the most important questions facing this country'. Table 4.19 shows changes in self-assessment through the campaign. Similarly to the increase in political interest presented previously, women on average reported the largest gains by Wave 3 (column *M∆*), but beyond that, there are no clear patterns of treatment effects: token and critical mass averages are somewhat smaller than men only and parity averages, although an average gain in 'women majority' races does stand out more. Dispersion

Table 4.19: Descriptive statistics: 'Subjective' political knowledge, BES '15 and Survey experiment

Gender	Treatm.	M Wave 1 BES	M Wave 3 BES	MΔ	M Round 1 Experiment	M Round 2 Experiment
Men	M.only	3.98	4.01	0.01	4.07	4.03
	Token	3.97	4.01	0.02	3.88	4.19
	C.mass	4.01	4.02	0.01	3.67	3.94
	Parity	4.00	4.04	0.03	3.58	4.06
	W.maj.	4.00	4.03	0.04		
Women	M.only	3.57	3.68	0.09	3.57	3.93
	Token	3.58	3.68	0.07	3.82	3.84
	C.mass	3.56	3.66	0.07	3.57	3.46
	Parity	3.57	3.69	0.09	3.53	3.86
	W.maj.	3.59	3.73	0.12		

differs across men and women but remain within a narrow range across treatment groups ($SD = 76$–79 and $SD = 82$–89, respectively).

I repeated this question in the survey experiment. The simple group averages in Table 4.19 show that the second round of experiments was more successful in reproducing the gender gap (perhaps due to this being asked after a more difficult political knowledge question), but that group averages across treatment groups are somewhat inconsistent with either positive or negative role model effects in either rounds. In the first round, there may have been a token spike on women's subjective assessment of their political knowledge, but a negative critical mass effect dominated in the replication round. Men's subjective assessment has gradually decreased as the proportion of women in picture stimuli increased in Round 1, but not in Round 2. There are no consistent patterns in dispersion across genders. The full range of standard deviations is $SD = 0.64$–1.00, critical mass driving the largest variability but I detected no further patterns beyond that.

Tables 4.20 and 4.21 report the model results for subjective political knowledge. Both dependent variables are standardised for better comparability with the other measures in this chapter. In the BES, constituency-level variation is estimated at 0, suggesting that substantively significant effects on this level, namely that of the proportion of women in campaigns, are unlikely. None of the two sets of models show a decrease in residual variances by a large magnitude. After inclusion of gender and treatment, explanatory power may be up to 2% in the survey experiment but less than 1% in the BES data. Nevertheless, these small variances are likely measurement artefacts, trying to teasing out variability on a single-item measure with few scale points. I aim to correct for this with my questions developed to measure confidence below.

Looking at the fixed effects in the BES models, women are found to have assessed their political knowledge more positively by the post-election wave, if only by a tenth of a standard deviation. On top of this, treatment effects are estimated to be near zero. Women respondents of the survey experiment report reliably, a quarter to a third standard deviation *worse* assessments of their political knowledge, compared with men—but no treatment effects are detected in this project either. Interestingly, posterior point predic-

Figure 4.6: Treatment x gender marginal effects, Models 10.2 & 11.2

Change in self-assessed political knowledge, BES Internet Panel 2015
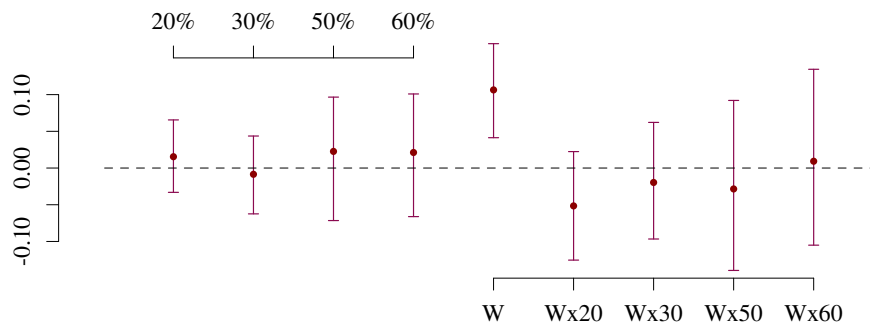
Self-assessed political knowledge, survey experiment
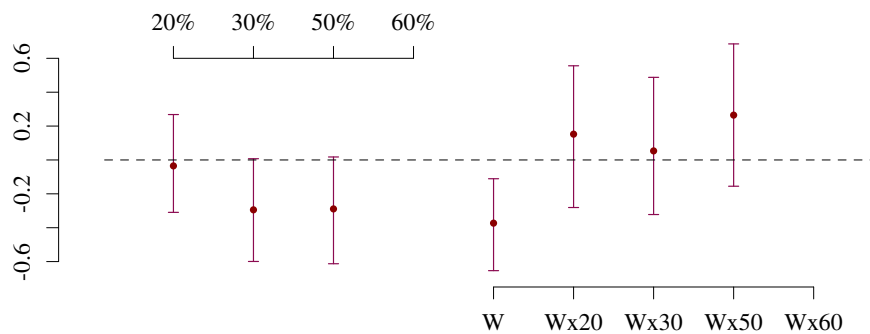
Key: β coefficients (posterior means) and 95% CI corresponding to the effect of women's presence in campaigns (BES) or picture stimuli (survey experiment) at the discrete proportions of near 20% or token, near 30% or critical mass, 50% or parity, and women's majority near 60%. W stands for women respondents.

tions do increase gradually, as shown by the marginal effects plot in Figure 4.6, but these effects are likely to counter-balance negative treatment effects on men. Again, at least in terms of the survey experiment results, these patterns may become sharper with a higher probability behind these effect sizes with a measurement approach that fixes problems with scale variability and uses multiple items.

Confidence

I am directly tackling concerns about scale variability and the number of response items by implementing a quiz question-level measure of confidence. After each question, as analysed under political knowledge, participants of the survey experiment were asked to indicate their level of confidence in the answer they were about to submit. Participants were allowed to guess an answer but asked explicitly to pick the minimum score on the 10-point scale, indicating a least confident answer. Table 4.22 shows group means across the two genders and treatment groups. In the first round of data collection, test questions were easier, thus the average confidence scores I report do not move much, except a moderate decrease in men's confidence gradually, from men only to parity exposure, and a very minimal increase of women's confidence by parity. These movements are much

Table 4.20: Models 10.0–10.2, Change in Subjective political knowledge, BES 2015

	m10.0	m10.1	m10.2
DV	*Subjective assessment of political knowledge,*		
	Wave 3–Wave1, standardised (z) scale		
Type	Linear / MCMC	Linear / MCMC	Linear / MCMC
Subject N	25748	25748	25748
Constituency N	632	632	632
DIC	73073.26	73042.8	73013.96
	Beta [95% CI]	Beta [95% CI]	Beta [95% CI]
Intercept	0.00[-0.01;0.01]	-0.05[-0.09;-0.01]	-0.04[-0.09;0.00]
Gender: Woman		0.11[0.05;0.17]	0.11[0.04;0.17]
Treatment: 20%		0.02[-0.03;0.07]	0.02[-0.03;0.07]
Treatment: 30%		-0.01[-0.06;0.05]	-0.01[-0.07;0.04]
Treatment: 50%		0.03[-0.06;0.11]	0.02[-0.07;0.10]
Treatment: 60%		0.02[-0.07;0.11]	0.02[-0.07;0.10]
Woman x 20%		-0.05[-0.12;0.02]	-0.05[-0.13;0.02]
Woman x 30%		-0.02[-0.10;0.06]	-0.02[-0.10;0.06]
Woman x 50%		-0.03[-0.15;0.08]	-0.03[-0.14;0.09]
Woman x 60%		0.01[-0.12;0.12]	0.01[-0.10;0.14]
Resp.: Age			-0.03[-0.05;-0.02]
Resp.: Educ.			0.01[0.00;0.02]
Const.: Marginality			0.00[-0.02;0.01]
	Gamma [95% CI]	Gamma [95% CI]	Gamma [95% CI]
Constit. var (Random)	0.00[0.00;0.00]	0.00[0.00;0.00]	0.00[0.00;0.00]
Residual var	1.00[0.98;1.02]	1.00[0.98;1.02]	1.00[0.98;1.01]

Table 4.21: Models 11.0–11.2, Subjective political knowledge, survey experiment

	m11.0	m11.1	m11.2
DV	*Subjective assessment of political knowledge,*		
	standardised (z) scale		
Type	Linear / MCMC	Linear / MCMC	Linear / MCMC
Subject N	605	605	605
Coll. round	2	2	2
	2	2	2
Platform N	2	2	2
DIC	1719.95	1711.6	1713.15
	Beta [95% CI]	Beta [95% CI]	Beta [95% CI]
Intercept	0.00[-0.08;0.08]	0.13[0.01;0.24]	0.26[0.04;0.48]
Gender: Woman		-0.26[-0.41;-0.11]	-0.37[-0.65;-0.11]
Treatment: 20%			-0.03[-0.32;0.26]
Treatment: 30%			-0.29[-0.56;0.04]
Treatment: 50%			-0.29[-0.60;0.03]
Woman x 20%			0.15[-0.22;0.60]
Woman x 30%			0.06[-0.35;0.46]
Woman x 50%			0.26[-0.16;0.68]
	Gamma [95% CI]	Gamma [95% CI]	Gamma [95% CI]
Coll.round var (Random)	0.00[0.00;0.00]	0.00[0.00;0.00]	12.14[0.00;0.01]
Crowds-pl. var (Random)	0.00[0.00;0.00]	0.05[0.00;0.00]	0.00[0.00;0.00]
Residual var	1.01[0.89;1.12]	0.99[0.89;1.11]	0.98[0.87;1.09]

sharper in the second round of data collection, where I implemented a difficult test, suggesting gradual role model effects on women's confidence and negative effects on men's confidence. Scrutinising descriptive results further, it is worth noting that the standardised scores to reflect group averages across both rounds point to a slight curvilinearity of effects. Men may have experienced a boost in confidence under women's token status, while women may have experienced a slight depression of confidence under the same exposure.

Table 4.22: Descriptive statistics: Confidence in test answers, survey experiment

Gender	Treatm.	*M* Round 1	*M* Round 2	*M* overall, *standard.* (z)
Men	M.only	3.23	7.13	0.22
	Token	3.23	7.83	0.37
	C.mass	3.21	7.33	0.25
	Parity	3.09	6.04	-0.14
Women	M.only	3.06	5.72	-0.24
	Token	3.06	5.68	-0.25
	C.mass	3.01	6.08	-0.21
	Parity	3.08	6.76	0.01

Models 12.0–12.2 tease out these effects on 6,708 responses nested within 605 test subjects and 22 test questions. Following the variance components or random effect 'Gammas' in Table 4.23, it becomes apparent that up to about 2 per cent of subject level variance may be explained by gender, while treatment may have very little explanatory power. Residual variances are insensitive to model expansion. Nevertheless, as far as fixed effects are concerned, my models detect evidence that the initial confidence gap between men and women is affected women's proportion in the picture stimuli. The average treatment effect by parity (on men) is a quarter standard deviation decrease in confidence, but this is countered by women's larger, third standard deviation increase in confidence (Beta 'Woman x 50%'), suggesting that exposure to a parity of women in picture stimuli may be enough to close the gender gap as observed under men only exposure. Figure 4.7 shows these marginal effects. Note that, given descriptive findings, these treatment effects are driven by those detected in the second round of survey experiment, and moderated by the results in the first round. Nevertheless, besides the difference in magnitude, the effects are consistent across the two rounds.

This concludes my discussion of one component of political efficacy, examined in isolation, namely confidence in knowledge and skills. I will discuss more about these findings in Section Conclusions 4.2.5. Next, I turn to observing another component of self-efficacy separately, namely perceived political influence, followed by analysis of an efficacy battery that includes components of subjective assessment, perceived influence of vote, and external political efficacy, that is, a judgement of whether political actors or institutions themselves have influence over policy in the first place.

Figure 4.7: Treatment x gender marginal effects, Model 12.2

Confidence in test answer, political knowledge quiz, survey experiment
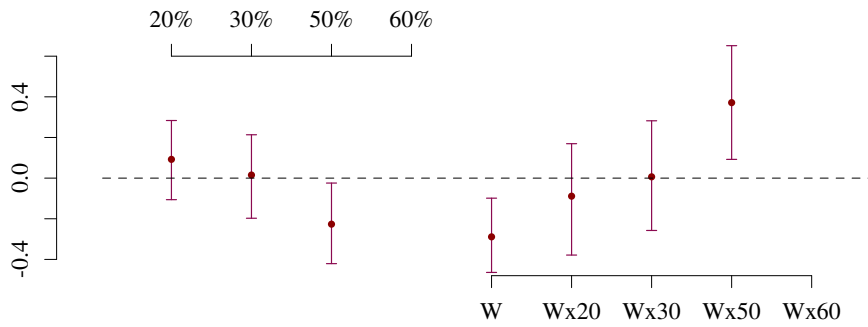
Key: β coefficients (posterior means) and 95% CI corresponding to the effect of women's presence in picture stimuli at the discrete proportions of near 20% or token, near 30% or critical mass, and 50% or parity. W stands for women respondents.

Table 4.23: Models 12.0–12.2, Confidence in political knowledge, survey experiment

	m12.0	m12.1	m12.2
DV	*Confidence in political knowledge, standardised (z) scale*		
Type	Linear / MCMC	Linear / MCMC	Linear / MCMC
Stacked N	6708	6708	6708
Test question N	22	22	22
Subject	605	605	605
Coll. round	2	2	2
Platform N	2	2	2
DIC	15409.72	15405.76	15405.18
	Beta [95% CI]	Beta [95% CI]	Beta [95% CI]
Intercept	-0.01[-0.17;0.20]	0.12[-0.08;0.30]	0.14[-0.07;0.34]
Gender: Woman		-0.23[-0.32;-0.13]	-0.29[-0.47;-0.11]
Treatment: 20%			0.09[-0.09;0.30]
Treatment: 30%			0.02[-0.19;0.22]
Treatment: 50%			-0.22[-0.41;-0.02]
Woman x 20%			-0.09[-0.37;0.18]
Woman x 30%			0.01[-0.26;0.27]
Woman x 50%			0.37[0.07;0.62]
	Gamma [95% CI]	Gamma [95% CI]	Gamma [95% CI]
Question var (Random)	0.15[0.07;0.25]	0.15[0.07;0.25]	0.14[0.07;0.25]
Subject var (Random)	0.35[0.30;0.39]	0.33[0.29;0.38]	0.33[0.29;0.38]
Coll.round var (Random)	0.01[0.00;0.00]	0.00[0.00;0.00]	0.00[0.00;0.00]
Crowds-pl. var (Random)	0.12[0.00;0.01]	0.05[0.00;0.00]	0.00[0.00;0.00]
Residual var	0.54[0.52;0.56]	0.54[0.52;0.55]	0.54[0.52;0.56]

Perceived influence

Perceived influence over the political outcomes, kept vague in question phrasing, is assessed using a single-item Likert scale in the 2010 BES panels but not repeated in 2015—latter used a multi-item battery of questions about political efficacy. I am thus able to test the working hypotheses that women became more efficacious, evidenced through 'perception of influence', in constituencies where more women ran for a seat. Table 4.24 shows the 2010 question means per gender and per proportion of women in campaigns.

Interestingly, respondents of both genders appear to have lost somewhat of their initial assessment, and men did more so than women. Women seem to have submitted only a slightly more pessimistic assessment of their influence in the first wave, and some of these differences evened out by wave 3. There is no clear pattern of effects regarding the proportion of women campaigning, but men seem to have become less pessimistic in races with a high percentage of women candidates.

Table 4.24: Descriptive statistics: Perceived political influence, scaled 1–5, BES 2010

Gender	Treatm.	M Wave 1	M Wave 3	$M\Delta$
Men	M.only	2.23	1.99	-0.20
	Token	2.33	2.02	-0.27
	C.mass	2.35	2.13	-0.27
	Parity	2.25	2.00	-0.14
	W.maj.	2.55	2.48	-0.07
Women	M.only	2.26	2.11	-0.12
	Token	2.12	1.99	-0.11
	C.mass	2.14	2.00	-0.11
	Parity	2.45	2.32	-0.20
	W.maj.	2.38	2.14	-0.18

The model-based answer to detect 'treatment' effects offers no clear patterns either. Table 4.25 shows that all marginal effects are estimated to be near zero, although the point estimates do suggest that women become less pessimistic (evidenced by a decreasing Wave 3 minus Wave1 differential) as the proportion of female candidates increased their constituencies: a movement invisible through group averages descriptively. The top panel of Figure 4.8 shows these effects. Finally, it is worth noting that only the inclusion of control variables on the individual and constituency levels drives down residual variance, and even that explains only about 1% variability in campaign effects on perceived political influence.

Political efficacy, multi-component

Political efficacy in the BES 2015 is assessed using multiple questions tapping into its different aspects, namely into voters' self-efficacy as active members of the citizenry, and into external political efficacy broadly assessing if citizens believe that policy makers are able and willing to implement good policy. At least three questions tapped directly self- or 'internal' type of efficacy, asking to provide a level of agreement with the following statements. Firstly, analysed above separately as subjective political knowledge, stating that respondent has 'a pretty good understanding of the important political issues facing our country'; secondly, that it 'takes too much time an effort to be active in politics and public affairs', and thirdly, that 'it is often difficult for [respondent] to understand what is going on in government and politics'. A fourth statement clearly taps into external political efficacy, that 'politicians' don't care what people like [respondent] think[s]'. I discarded further statements either because they were not repeated across the pre- and post-election waves to analyse campaign effects, or because they tap into some other dimension that I deemed less relevant for the current study. For example, while one

Table 4.25: Models 13.0–13.2, Change in political efficacy BES 2010 (single item)

	m13.0	m13.1	m13.2
DV	*Perceived political influence,*		
	Wave 3–Wave1, standardised (z) scale		
Type	Linear / MCMC	Linear / MCMC	Linear / MCMC
Subject N	10275	10275	10275
Constituency N	630	630	630
DIC	29162.33	29160.98	29134.26
	Beta [95% CI]	Beta [95% CI]	Beta [95% CI]
Intercept	0.00[-0.02;0.02]	-0.02[-0.07;0.04]	-0.01[-0.06;0.04]
Gender: Woman		0.06[-0.02;0.14]	0.05[-0.03;0.13]
Treatment: 20%		-0.03[-0.10;0.03]	-0.04[-0.10;0.03]
Treatment: 30%		-0.03[-0.10;0.06]	-0.03[-0.12;0.05]
Treatment: 50%		0.02[-0.10;0.15]	0.02[-0.12;0.14]
Treatment: 60%		0.08[-0.07;0.22]	0.06[-0.07;0.21]
Woman x 20%		0.03[-0.06;0.13]	0.04[-0.05;0.14]
Woman x 30%		0.01[-0.10;0.12]	0.02[-0.09;0.14]
Woman x 50%		-0.08[-0.25;0.11]	-0.08[-0.27;0.08]
Woman x 60%		-0.09[-0.28;0.13]	-0.07[-0.28;0.13]
Resp.: Age			-0.06[-0.07;-0.04]
Resp.: Educ.			0.01[-0.01;0.03]
Const.: Marginality			-0.01[-0.03;0.01]
	Gamma [95% CI]	Gamma [95% CI]	Gamma [95% CI]
Constit. var (Random)	0.00[0.00;0.00]	0.00[0.00;0.00]	0.00[0.00;0.00]
Residual var	1.00[0.97;1.03]	1.00[0.97;1.03]	0.99[0.97;1.02]

might consider whether respondent thinks voting for small parties is a waste of effort, or whether it matters if they vote for a party with a chance to win, but I found no clear conceptual link between these aspects of political efficacy and women's empowering or role model effect in visible roles.

Table 4.26 shows BES 2015 group means averaged across internal and external efficacy items. Typology matters: both men and women seem to judge their personal efficacy a little better during the campaign but judge the efficacy of politicians in office less optimistically after the campaign. Note that positive changes in internal efficacy are against the findings on the perceived influence scale, which is likely due to these items relying heavily on assessments of political knowledge rather than the explicitly on the influence component. Women's gains in internal efficacy seemingly surpass men's as the proportion of women increases in the campaign environment.

I implemented these four questions in both rounds of the survey experiment. Table 4.27 shows gender and treatment group means on these items. There are no major differences between the results from these rounds and point to small effects, generally in the direction that men's self-efficacy decreased by parity exposure, while women reported greater self-efficacy only at token and critical mass proportions but a decrease when exposed to a parity of male and female politicians, suggesting curvilinearity. In terms of external efficacy, effect directions across data collection rounds appear entirely inconsistent.

Table 4.26: Descriptive statistics: Political efficacy, BES 2015

Gender	Treatm.	Internal M w1	Internal M w3	External M w1	External M w3	Δ In.	Δ Ex.
Men	M.only	3.46	3.55	3.71	3.59	0.09	-0.12
	Token	3.48	3.54	3.66	3.59	0.06	-0.07
	C.mass	3.51	3.56	3.67	3.57	0.05	-0.10
	Parity	3.55	3.58	3.60	3.49	0.03	-0.11
	W.maj.	3.47	3.54	3.66	3.56	0.07	-0.10
Women	M.only	3.21	3.31	3.63	3.53	0.10	-0.10
	Token	3.21	3.33	3.65	3.56	0.12	-0.09
	C.mass	3.19	3.32	3.68	3.56	0.13	-0.12
	Parity	3.20	3.35	3.61	3.57	0.15	-0.04
	W.maj.	3.19	3.32	3.67	3.60	0.13	-0.07

Table 4.27: Descriptive statistics: Political efficacy, Survey experiment

Gender	Treatm.	Internal M Round 1	Internal M Round 2	External M Round 1	External M Round 2
Men	M.only	3.52	3.36	2.17	3.89
	Token	3.37	3.22	2.49	3.39
	C.mass	3.39	3.33	2.52	3.49
	Parity	3.24	3.26	2.34	3.68
Women	M.only	3.25	3.30	2.51	3.54
	Token	3.22	3.48	2.33	3.45
	C.mass	3.19	3.42	2.34	3.89
	Parity	3.11	3.39	2.31	3.83

My models explore effects on political efficacy as a single latent construct. Models 14.0–14.2 on campaign effects point to low variance components on the constituency level casting doubt on the explanatory power of candidate gender. Only subject level variation seems significant, explaining about 5 per cent of variation in changes on the standardised scale. Nevertheless, fixed effects point to treatment effects on women. Visualised in the middle panel of Figure 4.8, there is enough evidence that the proportion of women as candidates in the 2015 parliamentary constituencies gradually improved women's political efficacy, likely driven by the self-efficacy items in the latent construct. Men were less efficacious with more women in campaigns. However, both effects break by women's parity presence, where gains up to parity seemingly diminish.

By contrast, these effects are not mirrored in my survey experiment. There is weak evidence behind a negative effect of token and critical mass proportions but the effect sizes are insufficient to change political efficacy as predicted in the baseline category: men. The bottom panel of Figure 4.8 shows these marginal effects. Finally, model expansion did not affect residual variance suggesting very poor explanatory power.

A follow-up analysis confirms that, if fitted separately on external efficacy, efficacy effects were driven by internal political efficacy or self-efficacy. Using campaign effects on external efficacy as the dependent variable, all treatment coefficients remain close to zero, the gender x treatment interactions ranging between $\beta = 0.01$–0.04. The only exception is a small effect under tokenism on men's external political efficacy, which has

decreased marginally by -0.05 in terms of standard deviation unit[19]. Looking at external political efficacy in the survey experiment, where Round 1 and Round 2 averages pointed in inconsistent directions, the model-based answer is seemingly driven by Round 1 results in that it detects positive effects on men's external efficacy, concentrated under tokenism, $\beta = 0.35$ in standard deviation unit[20]; and negative effects on women's external efficacy, concentrated under critical mass, $\beta = -0.53$ in standard deviation unit[21].

Figure 4.8: Treatment x gender marginal effects, Models 13.2 and 14.2
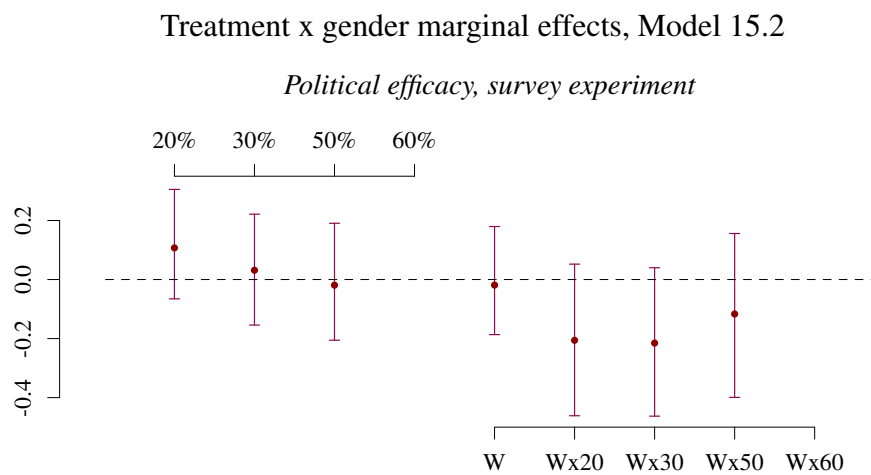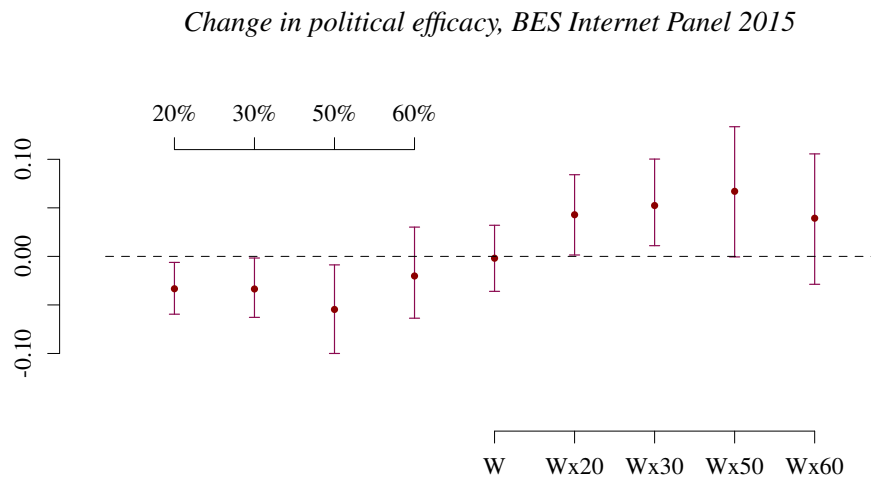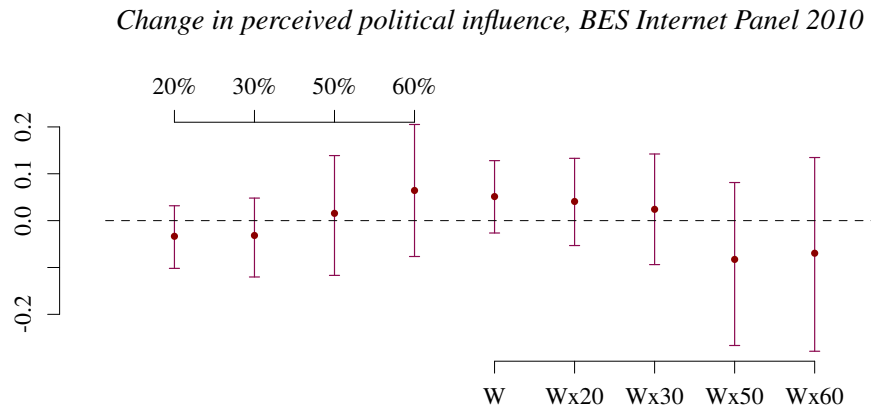
Change in perceived political influence, BES Internet Panel 2010

Change in political efficacy, BES Internet Panel 2015

Treatment x gender marginal effects, Model 15.2

Political efficacy, survey experiment

Key: β coefficients (posterior means) and 95% CI corresponding to the effect of women's presence in picture stimuli (survey experiment) at the discrete proportions of near 20% or token, near 30% or critical mass, and 50% or parity. W stands for women respondents.

Table 4.28: Models 14.0–14.2, Change in political efficacy, BES 2015 (multi item)

	m14.0	m14.1	m14.2
DV	*Political efficacy, Wave 3–Wave1, standardised (z) scale*		
Type	Linear / MCMC	Linear / MCMC	Linear / MCMC
Stacked N	95000	95000	95000
Item N	4	4	4
Subject	23750	23750	23750
Constituency N	632	632	632
DIC	268825.48	268805.25	268789.94
	Beta [95% CI]	Beta [95% CI]	Beta [95% CI]
Intercept	0.00[-0.01;0.01]	0.01[-0.02;0.03]	0.01[-0.01;0.03]
Gender: Woman		0.00[-0.03;0.04]	0.00[-0.03;0.03]
Treatment: 20%		-0.03[-0.06;0.00]	-0.03[-0.06;0.00]
Treatment: 30%		-0.03[-0.06;0.00]	-0.03[-0.06;0.00]
Treatment: 50%		-0.05[-0.10;-0.01]	-0.05[-0.10;-0.01]
Treatment: 60%		-0.02[-0.06;0.03]	-0.02[-0.06;0.03]
Woman x 20%		0.04[0.01;0.08]	0.04[0.00;0.08]
Woman x 30%		0.05[0.01;0.10]	0.05[0.01;0.10]
Woman x 50%		0.07[0.00;0.13]	0.07[0.00;0.13]
Woman x 60%		0.04[-0.03;0.10]	0.04[-0.02;0.11]
Resp.: Age			-0.01[-0.02;-0.01]
Resp.: Educ.			0.00[0.00;0.01]
Const.: Marginality			0.00[-0.01;0.00]
	Gamma [95% CI]	Gamma [95% CI]	Gamma [95% CI]
Item var (Random)	0.00[0.00;0.00]	0.00[0.00;0.00]	0.00[0.00;0.00]
Subject var (Random)	0.05[0.04;0.06]	0.05[0.04;0.06]	0.05[0.04;0.06]
Constit. var (Random)	0.00[0.00;0.00]	0.00[0.00;0.00]	0.00[0.00;0.00]
Residual var	0.95[0.94;0.96]	0.95[0.94;0.96]	0.95[0.94;0.96]

Table 4.29: Models 15.0–15.2, Political efficacy (multi item), survey experiment

	m15.0	m15.1	m15.2
DV	*Political efficacy, standardised (z) scale*		
Type	Linear / MCMC	Linear / MCMC	Linear / MCMC
Stacked N	2420	2420	2420
Item N	4	4	4
Subject	605	605	605
Coll. round	2	2	2
Platform N	2	2	2
DIC	5944.99	5941.17	5944.35
	Beta [95% CI]	Beta [95% CI]	Beta [95% CI]
Intercept	-0.02[-0.88;0.86]	0.06[-0.89;0.98]	0.00[-1.00;0.89]
Gender: Woman		-0.15[-0.23;-0.05]	-0.02[-0.20;0.16]
Treatment: 20%			0.11[-0.07;0.30]
Treatment: 30%			0.03[-0.15;0.23]
Treatment: 50%			-0.02[-0.21;0.19]
Woman x 20%			-0.21[-0.44;0.06]
Woman x 30%			-0.22[-0.45;0.04]
Woman x 50%			-0.12[-0.38;0.16]
	Gamma [95% CI]	Gamma [95% CI]	Gamma [95% CI]
Question var (Random)	0.72[0.05;2.62]	0.61[0.03;1.74]	4.58[0.05;2.61]
Subject var (Random)	0.22[0.19;0.27]	0.22[0.18;0.26]	0.22[0.18;0.26]
Coll.round var (Random)	0.00[0.00;0.00]	0.00[0.00;0.00]	0.00[0.00;0.00]
Crowds-pl. var (Random)	0.02[0.00;0.00]	0.00[0.00;0.00]	0.28[0.00;0.01]
Residual var	0.59[0.55;0.63]	0.59[0.55;0.62]	0.59[0.55;0.63]

4.2.4 Trust

Trust in MPs is measured in a question battery together with trust in a number of other people and institutions. The same Likert scales were implemented in both election years in the BES. Tables 4.30 and 4.31 show group averages through the 2010 waves: men and women both grew more trusting of MPs by the end of both election periods, Delta M column showing within-individual average changes. Regarding the proportion of women running, there are baseline differences in 2010 showing that citizens were more trusting of politicians in constituencies where women eventually ran for seats. The magnitude of change, in terms of campaign effects on trust, may still probe to be dependent on the proportion of female candidates in campaigns: in 2010, men grew less trusting of politicians in constituencies with a growing presence of women as candidates while women clearly became more trusting. Discarding women majority constituencies, it seems that this pattern is repeated in 2015 if much more moderate.

Comprable across data collection modes, Table 4.32 shows standard scores of average trust per groups together with the survey experiment. Group averages in latter are inconsistent with campaign effects in the BES: men and women were most trusting of MPs under token exposure, while both groups less trusting as the proportion of women increased beyond token in picture stimuli—suggesting this be a kind of tokenism that uniformly affects men and women, unlike my results regarding political interest and learning in the BES.

Table 4.30: Descriptive statistics: Trust in MP's, BES 2010

Gender	Treatm.	M Wave 1	M Wave 3	$M\Delta$
Men	M.only	3.26	3.98	0.65
	Token	3.53	4.13	0.64
	C.mass	3.91	4.34	0.38
	Parity	3.67	3.97	0.03
	W.maj.	4.43	4.25	-0.08
Women	M.only	3.46	4.08	0.57
	Token	3.59	4.21	0.65
	C.mass	3.60	4.14	0.48
	Parity	3.06	3.82	0.74
	W.maj.	4.18	4.72	1.00

Table 4.31: Descriptive statistics: Trust in MP's, BES 2015

Gender	Treatm.	M Wave 1	M Wave 3	$M\Delta$
Men	M.only	3.18	3.41	0.22
	Token	3.23	3.38	0.16
	C.mass	3.23	3.39	0.17
	Parity	3.26	3.43	0.17
	W.maj.	3.22	3.45	0.23
Women	M.only	3.17	3.36	0.19
	Token	3.20	3.36	0.16
	C.mass	3.20	3.36	0.14
	Parity	3.11	3.33	0.23
	W.maj.	3.16	3.26	0.13

Table 4.32: Descriptive statistics: Trust in MP's, all data, standard scores (*z*)

Gender	Treatm.	*M* BES '10	*M* BES '15	*M* Exp. Round 1	*M* Exp. Round 2
Men	M.only	-0.08	0.02	-0.15	-0.34
	Token	-0.01	0.01	0.29	0.37
	C.mass	0.09	0.01	-0.12	0.00
	Parity	-0.09	0.04	0.12	-0.13
	W.maj.	0.05	0.05		
Women	M.only	-0.04	-0.01	-0.02	0.05
	Token	0.03	-0.01	0.23	0.23
	C.mass	-0.01	-0.01	-0.28	0.02
	Parity	-0.16	-0.03	-0.03	-0.18
	W.maj.	0.27	-0.07		

The model tables show that the variance explained by my host of predictors is negligible, except in the survey experiment where inclusion of treatment does improve about 2% of model fit, likely because of tokenism. In both approaches, the token fixed effects are largest, affecting men and women uniformly but in inconsistent directions. In the BES, people grew more trusting of MPs but the magnitude of this change is *smallest* in token constituencies; whereas in the survey experiment, trust peaks for both men (the baseline Beta, Treatment: 20%) and women (Beta 'Woman x 20%" negative but very large posterior variance, making it unlikely to reverse baseline tokenism). Figure 4.9 shows these marginal effects.

Table 4.33: Models 16.0–16.2, Change in trust in MPs, BES 2010&2015

	m16.0	m16.1	m16.2
DV	*Trust in MPs, Wave 3–Wave1, standardised (z) scale*		
Type	Linear / MCMC	Linear / MCMC	Linear / MCMC
N	27022	27022	27022
Constituency N	632	632	632
Election years	2	2	2
DIC	76314.54	76323.65	76281
	Beta [95% CI]	Beta [95% CI]	Beta [95% CI]
Intercept	0.00[-0.02;0.01]	0.04[0.00;0.08]	0.04[-0.01;0.08]
Gender: Woman		-0.02[-0.07;0.04]	-0.01[-0.07;0.04]
Treatment: 20%		-0.05[-0.10;0.00]	-0.05[-0.10;0.00]
Treatment: 30%		-0.04[-0.10;0.01]	-0.04[-0.10;0.01]
Treatment: 50%		-0.05[-0.14;0.03]	-0.05[-0.15;0.03]
Treatment: 60%		-0.01[-0.10;0.07]	-0.01[-0.10;0.07]
Woman x 20%		0.01[-0.05;0.07]	0.02[-0.04;0.09]
Woman x 30%		-0.01[-0.09;0.06]	0.00[-0.08;0.07]
Woman x 50%		0.07[-0.05;0.18]	0.07[-0.04;0.20]
Woman x 60%		-0.03[-0.15;0.08]	-0.03[-0.15;0.08]
Resp.: Age			0.04[0.03;0.05]
Resp.: Educ.			0.02[0.00;0.03]
Const.: Marginality			0.00[-0.01;0.01]
	Gamma [95% CI]	Gamma [95% CI]	Gamma [95% CI]
Constit. var (Random)	0.00[0.00;0.01]	0.00[0.00;0.01]	0.00[0.00;0.01]
Election year var (Random)	0.00[0.00;0.00]	0.00[0.00;0.00]	0.00[0.00;0.00]
Residual var	0.98[0.97;1.00]	0.98[0.97;1.00]	0.98[0.97;1.00]

Table 4.34: Models 17.0–17.2, Trust in MPs, survey experiment

	m17.0	m17.1	m17.2
DV	*Trust in MPs, standardised (z) scale*		
Type	Linear / MCMC	Linear / MCMC	Linear / MCMC
Subject N	605	605	605
Coll. round	2	2	2
Platform N	2	2	2
DIC	1718.98	1720.97	1712.8
	Beta [95% CI]	Beta [95% CI]	Beta [95% CI]
Intercept	-12.36[-0.08;0.09]	0.01[-0.11;0.12]	-0.24[-0.45;-0.01]
Gender: Woman		-0.02[-0.18;0.14]	0.26[-0.08;0.55]
Treatment: 20%			0.57[0.26;0.88]
Treatment: 30%			0.18[-0.15;0.51]
Treatment: 50%			0.24[-0.06;0.58]
Woman x 20%			-0.37[-0.80;0.06]
Woman x 30%			-0.36[-0.77;0.11]
Woman x 50%			-0.37[-0.79;0.09]
	Gamma [95% CI]	Gamma [95% CI]	Gamma [95% CI]
Coll.round var (Random)	14.89[0.00;0.00]	0.00[0.00;0.00]	0.00[0.00;0.00]
Crowds-pl. var (Random)	0.00[0.00;0.00]	0.00[0.00;0.00]	0.00[0.00;0.00]
Residual var	1.00[0.89;1.11]	1.00[0.89;1.11]	0.98[0.86;1.09]

4.2.5 Section conclusions

In this section I presented results from an observation study and an experimental study side by side, grouped by dependent variable, to investigate if evidence is consistent that exposure to female candidates in increasing proportions improves psychological engagement with politics. While I will look at broad patterns and groups of effects across both political engagement and attentional engagement in the Chapter Conclusions (next section), here I am briefly summarising my results, focusing on consistency across observational and experimental approaches.

Women's political interest is impacted by tokenism in the BES and only one round of survey experiments. Learning about politics follows this pattern in the BES, but it seems that both men's and women's performance on the political knowledge test asked in the survey experiment is improved by token exposure. Women assess their subjective political knowledge more pessimistically than men do in both study types, and the BES study finds that women became more optimistic about their subjective political knowledge through the campaign. These movements are, however, not impacted by women's proportion in campaigns, and evidence is weak that subjective political knowledge increases as more female candidates are shown in picture stimuli. Confidence about political knowledge, measured in the survey experiment after each test question, does shows gradual improvements, and women became very confident about their test answers by parity. This effect is not mirrored on perceived political influence but it is detected on a more comprehensive set of questions about political efficacy: women became more efficacious in constituencies where more women ran in the 2015 election. This is not replicated in the survey ex-

Figure 4.9: Treatment x gender marginal effects, Models 16.2 & 17.2
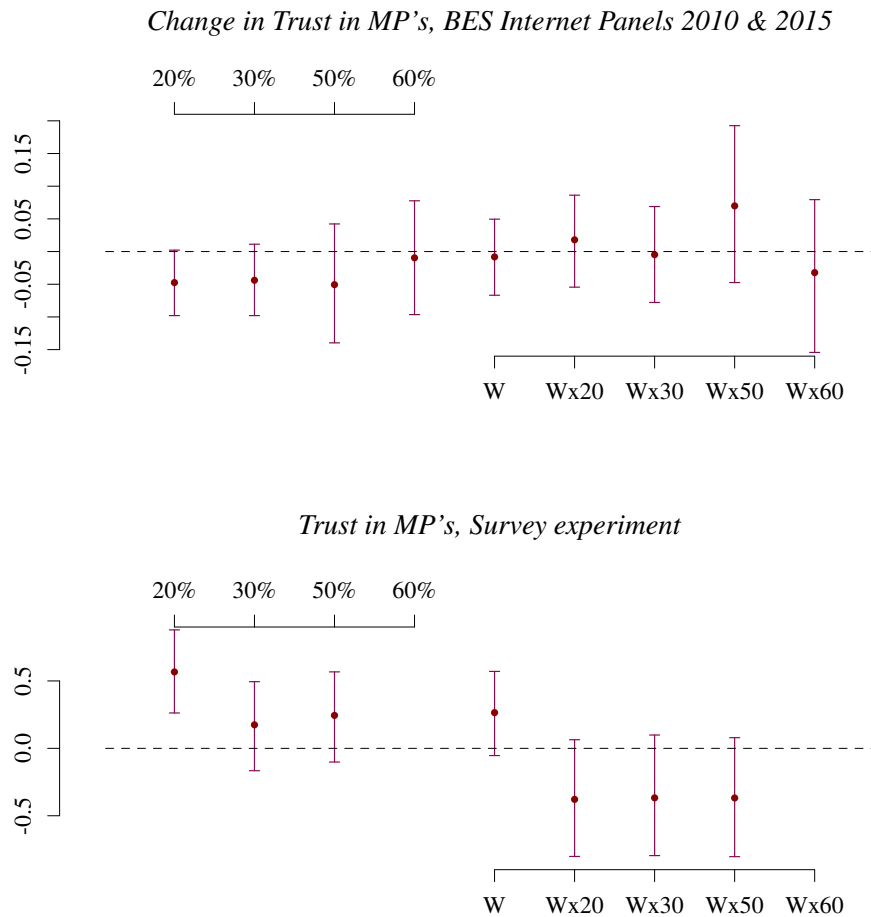
Change in Trust in MP's, BES Internet Panels 2010 & 2015

Trust in MP's, Survey experiment

Key: β coefficients (posterior means) and 95% CI corresponding to the effect of women's presence in campaigns (BES) or picture stimuli (survey experiment) at the discrete proportions of near 20% or token, near 30% or critical mass, 50% or parity, and women's majority near 60%. W stands for women respondents.

periment using similar questions. Looking at external efficacy separately as a dependent variable, however, both the BES and the survey experiment models suggested positive token effects on men's, suggesting that they judged politics itself to be more responsive and responsible with minimal presence of women. Finally, trust is impacted across the BES and the survey experiments in inconsistent directions.

4.3 Chapter Conclusions

In this chapter, I reviewed observational and experimental evidence in support of role model effects on a number of potential outcomes, to explore if an increase in women's numerical representation increases women's 'attentional engagement' with politics, their self-assessed political interest, political knowledge (performance and confidence), political efficacy and trust in politicians. Across these dependent measures, two noteworthy patterns emerged: a non-linear token effect on attention-related variables, and a linear or gradual effect on self-efficacy related variables.

Token spike

Looking at psychophysiological data, as well as repeated survey measures from the British Election Study panels, I found that, although lower numbers of women pay attention to politics than men, regardless of whether featuring only men or a high proportion of women politicians, there is a spike in attention when presented with a single or token female politician. In the lab experiment, a cognitive test and a host of eye-tracking measures showed that women's token presence in pictures of political groups drew participants' attention, but in the few cases when their proportion increased and surpassed men's, this attention surplus diminished.

Correspondingly, women panel respondents in the 2010 and 2015 BES became more interested in the election than men by the end of both campaign periods but on top of this mechanism, in constituencies where campaigns featured a token female candidate, women became most interested. Perhaps stemming from a surplus in attention, my models using BES data also showed that the odds of learning about politics were highest in these token constituencies. Importantly, however, the survey experiment using randomised picture exposure offered mixed evidence. While women were most interested in politics under token exposure in the first round of data collection confirming tokenism, they were unaffected in the replication round.

Although a token spike has not yet been identified in the previous studies, the observation does link to a stream of literature on symbolic representation about the *novelty* of women on the campaign trail. In a long-standing empirical project exploring role model effects on political discussion, a recent US panel study [21] provides evidence that women were likely to discuss more about politics as a result of 'unique and newsworthy' [21, p.424] female candidates, where 'novelty' means that the candidate herself had not previously ran for office, and the state is such that it had not already been represented by a woman in the House, the Senate, or as Governor. A recent survey of political engagement finds some evidence that mixed-gender races, that is when a woman is standing against a man, increase political interest, decrease alienation and 'indifference' of candidates, but not same-gender races including a woman against a woman [129].

My results provide further evidence that variables related to attention to politics, and attention's potential consequences such as the frequency of political discussion or political knowledge of current affairs, are influenced by women's presence *as long as it carries a novelty factor*, if there is a sharp enough contrast to draw attention. It seems that in the context of the 2010 and 2015 elections, one woman challenging men may have invoked more election interest than a second or third woman entering the campaign in which case average election interest is 'back' to the figures when no women at all ran.

My experimental evidence of tokenism, however, is only conclusive on the micro-level, in the lab experiment. In the online experiment, I did not succeed in replicating my own finding to causally link exposure to a token woman to higher self-assessed political interest. The null result is, naturally, open to speculation on many points: about mea-

surement (single Likert item to self-report) in the context of the small N in the individual study rounds, and that the effect sizes in question are very small both in the lab and in the BES. I would not, however, discard substantive reasons behind the null effects. On the micro level, a token woman presence may invoke immediate interest in the stimulus but experimental subjects may not have consciously translated this into *political* interest. In the BES, while the survey question is similar, exposure to token female politicians is repetitive, spread out through the entire campaign period, and through multiple channels including media appearance and campaign material. It is this component of *repetitive exposure*, and to a variety of stimuli, is what future experimental investigation may need to tackle.

Linear role model effects

Unlike the 'token spike', a more gradual pattern emerged on some of the dependent variables, centering on self-efficacy. In the survey experiment, women were more and more confident about their political knowledge, interpreted as part of political efficacy [48],[49], as the proportion of women increased in the pool of pictures they viewed, and gave most confident answers when presented with a parity of men and women. In terms of broader measures of political efficacy including perceptions about self-assessed political influence, no such pattern emerged in the survey experiment but it did in the 2015 BES: women, taking their baseline efficacy prior to the campaign, became more and more efficacious as the proportion of women campaigning in their constituency increased. Separating the concepts of self- or internal efficacy about one's own political influence from trust in MPs and external efficacy, my models found that former is implicated while latter may not be.

This linear role model effect, namely, that a gradually improving numerical balance gradually increases political engagement, is seemingly not driven by a novelty factor. Although there were not many Critical mass and parity pictures in the pool of visuals in the lab experiment, the first evidence does suggest that attention was not biased to these proportions as much as they were to token pictures. In other words, something *beyond mere attention surplus* could be at work when looking at effects on self-efficacy. The following chapters test possible channels of mediating mechanisms explicitly, including effects through alleviating anxiety and stereotype threats, which may be the explanation behind these linear effects. In another chapter, I test mediation through improving expectations about policy output, which may prove to be influential self-efficacy judgments as well.

At the core of the argument, the role model hypothesis was simply that if more and more women participate in politics in visible roles, more and more women in the electorate will engage with politics. My analysis shows different indicators of psychological engagement are impacted differently by gender balance, and this schematic form of the role model hypothesis is most likely to apply on self-efficacy, that is women's confidence in their competency and potency in the political world.

Further considerations

In this chapter, I pooled together analysis of a host of indicators measuring psychological engagement with politics to get a comprehensive review of role model effects, with previous research testing only in isolation or reporting variables selectively. I conclude that the majority of evidence altogether supports some form of the role model hypothesis, while the pooled evidence from two rounds of survey experiments yielded inconclusive effects, potentially casting doubt on a simplistic causal link between picture exposure and self-reports of political engagement. I did not address some of the *negative effects* found in the replication round of the survey experiment, such as a seemingly negative impact of critical mass exposure on political efficacy, or negative token and critical mass effects on trust in MPs. Although negative effects would be consistent with the self-deflating or 'harmful social comparison' theories briefly mentioned in Chapter 2 Section 2.1.2, there are three main reasons I do not see these as direct evidence: inconsistency across dependent measures, inconsistency across data collection rounds, and inconsistency with the lab experiment and the observational study.

Although conceptually separate, there is a degree of empirical overlap between self-reported political interest, subjective political knowledge, political efficacy, and trust. It is likely that a negative effect would impact most of these consistently, which is not the case: women assessed their political knowledge better when viewed pictures with at least a token woman in them, and the pattern is somewhat similar, if more variable, for political interest. Second, the negative critical mass tendency is mostly driven by the replication round of data collection, these effects were entirely missing from the original data collection round, showing larger than expected volatility of the survey responses. Third, there is no corresponding tendency either in the lab experiment regarding attentional bias, or in the BES data that would lend more support to self-deflating social comparisons, which were both more supportive of role model effects. These considerations do not mean that the negative effects should be entirely discarded from my analyses, but are meant to explain why I choose not to engage with these results on a more theoretical level.

Finally, I did not interpet treatment effects on the outgroup members explicitly. Implicitly, men are always in the analysis because I reserved the term of 'role model effects' for cases when they were unique to women. This is formalised in the analysis with the interaction term: a significant gender and treatment interaction means in practice that men were either unaffected, or negatively affected by the treatment. Both movements can be, however, of substantive interest. Interstingly, men were less likely to learn about foreign leaders in token constituencies, assessed their political knowledge lowest if critical mass or parity group respondents of the survey experiment, and were least confident in their test answers in the parity group—one might speculate if this is because of perceptions about group hierarchy and changes in dominance structures. Looking at self-confidence, men may have experienced a *token boost* in confidence, another aspect of tokenism that may be worth exploring in furture studies, especially in relation to Stereotype Boost Theory

[156],[157].

Notes

[5]These would be connected by arbitrarily small saccadic movements, which in reality took 20-30 milliseconds, see Table 4.1 on page 46

[6]My special thanks go to Ekaterina Kolpinskaya, looking through media content of the near one hundred twitter handles I selected for inspection.

[7]This is for purposes of exploratory and future analyses. Political engagement data might be linked directly to attentional engagement patterns, however, it is outside the scope of the current studies.

[8]The BES website itself has major party candidate data, but that is insufficient to determine the proportion of women amongst all candidates.

[9]The variable is standardised by centering on its mean and dividing by standard deviation. Thus a value of 0 in the model table means $0 \cdot SD_X + M_X = 0 \cdot 94.41 + 393.62 = 393.62$

[10]I excluded practice trials.

[11]I excluded incorrect responses from further analysis

[12]I give more detailed explanation on the appropriateness of multilevel modelling in Chapter 3, as well as a note at the beginning of this Chapter.

[13]Note that this is not pre-set for any participant but pairs appeared in randomised order.

[14]For this analysis, both rounds targeted 300 respondents, see Methodology Chapter. In the first round, I encountered difficulties recruiting women participants through the CrowdFlower platform, and I thus needed this to be on the field for longer. The second round of data collection targeted a sample of 750 but 2/3s of the sample received further manipulation, to be explored in the subsequent chapters.

[15]Most often non-English names as well as those names where predicted gender probability was too close to call.

[16]These numbers exclude Northern Ireland constituencies as the BES does not sample respondents from that country.

[17]The choice of using Bayesian modelling over Maximum Likelihood is due to some of the models, especially the latent ones, having a complicated nesting structure that would have convergence problems with the non-Bayesian packages available currently. In an earlier draft of this chapter, many of these models were fit by Maximum Likelihood, producing identical results.

[18]Though the component may not have reached convergence, see diagnostic plots

[19]95% credible intervals -0.11– -0.01.

[20]95% credible intervals 0.06–0.60.

[21]95% credible intervals -0.93– -0.15.

Chapter 5

Understanding the role model effect:
Stereotype threat and affect as mediating mechanisms

[...] We believe that it would be productive to look once again at the messages received by males and females of various ages in order to understand how the taste for politics emerges. [130]

[...] Even if a young woman emerges from childhood as a relatively cool and resilient adult, she still has to do battle with social forces that seem bent on making her anxious. [158]

My studies report evidence that, when present in increasing proportions, women in politics alleviate situation-specific anxieties when it comes to assessing one's own political skills; that at least a single standing 'token' woman may be relied on if one does not feel that taking action to confront stereotypes is possible; but that in the presence of anger, women confronted stereotypes, and did not turn away from politics.

As an important argument behind descriptive representation, studies have described a number of 'symbolic' or non-policy benefits that seem to generate directly from (women's) greater political representation, often described broadly as representatives instilling a 'sense of empowerment' [22],[30],[159]. While an uncertain claim in terms of tangible effects, a number of studies, including my thesis, takes this to suggest that increasing descriptive representation can foster political engagement, and thus more women in visible political office could increase women's participation, interest, and efficacy in politics.

There is, however, further uncertainty about the underlying *mediating mechanisms* as to why descriptive representation leads to increased political engagement. Only very few empirical studies make attempts to demonstrate a more thorough causal mechanism that is convincing on the individual level, one of the hypotheses being that women may expect their interests be better represented by women politicians [22],[19]. On this level of voter psychology, as I argue in Chapter 2 especially in Sections 2.1 and 2.1.2, there have been no competing hypotheses tested in the representation literature.

In the previous chapter, of 'Measuring role model effects', I only tested 'blank' causation between treatment and effect, namely exposure to an increasing number of women politicians on a range of indicators of attentional engagement first, and then on psychological engagement with politics. I concluded that (a) women's 'striking minority' or token presence may have the greatest impact on attention-related outcomes such as election interest or learning about politics, and (b) a more gradual increase in women's presence towards parity is most likely to affect deeper psychological variables such as confidence in political knowledge and self-efficacy in general.

In this chapter, I test hypotheses derived directly or indirectly from the stereotype threat paradigm predicting that women's attention to, and self-efficacy in politics is increased at least partly because, as role models, visible female politicians *help alleviate often implicit anxieties* stemming from negative stereotypes about women's role and capabilities in politics. For the purposes of this chapter, I call the process implicit because it does not presuppose (nor does it exclude) that women more instrumentally consider if their own policy preferences are being advanced by fellow women in office, an option tested in more detail in Chapter 6. Rather, anxieties may guide behaviour automatically, generating from subtle signals from the political environment, such as an evaluative apprehension in the presence of out-group members, namely men, advantaged by the same gender stereotypes that harm women in the political domain. I elaborate more on the idea of stereotype threat as situational performance anxiety in Chapter 2.

While ideally I would test mediation on all variables tested in Chapter 4 as well, I have to accept data limitations. Following experimental procedures I expand on below, I am able to test for mediation mainly in the survey and the lab experiment data sets, which also means that I am tied to effects observed in those data sets. The survey experiment was, generally speaking, unsuccessful in detecting effects on most of my survey measures, except *confidence about test answers*. Therefore, I can use confidence as the dependent variable in two of the following mediation studies, treating it as an example of a gradual role-model effect on self-efficacy[22]. For the token spike-type of effect on *attention*, I am able to test some mediation at the end of this chapter using the lab experiment data, but unable to scrutinise tokenism on political interest, an effect detected in the BES only but inconclusive across two rounds of survey experiments.

My contribution here is twofold. Firstly, as argued in Chapter 2, the stereotype threat literature has been especially useful in demonstrating causal linkage between experimentally induced stereotypes (very explicit reminders of stereotypes), anxiety, and malperformance on tasks; however, the experimental paradigm itself has difficulties when applied to study naturally and more subtly occurring stereotypes such as those in political representation. In the lab, stereotypes are 'written out' explicitly, whereas outside of the lab, stereotypes are 'in the air'[41]. In my survey experiment, introduced in Chapter 4, treatment is not artificial induction of threat but threat levels are a function of gender balance in picture stimuli, with the primary hypothesis being that increasing exposure to pictures of candidates as role models alleviate stereotype threats. This methodology is meant to

reflect more how stereotype threat in the political domain occurs in the real world such as through media exposure or perhaps even on the ballot paper.

Secondly, stereotype-induced anxiety has been linked extensively to testing performance, especially on mathematics tests [32],[37],[160],[161], with somewhat limited scope and only very few studies go beyond standard tests to implicate more general concepts such as 'intellectual identity' [41] or 'academic disidentification' [42]. Mine is an attempt to link stereotypes to self-efficacy and confidence about political knowledge which manifest in a much less-defined set of behaviours, but may ultimately result in engagement with, or disengagement from, politics.

Through the studies reported here, I am taking a range analytical and measurement strategies to get at the role of affect, predominantly anxiety, as mediator of role model effects. *Study 1*, the simplest experimental approach, follows the logic of the survey experiment first reported in Chapter 4, in that participants are shown pictures of groups of politicians with the proportion of women in them experimentally manipulated. After treatment delivery, they provided a quick assessment either of state anxiety (a general assessment of emotion felt 'right now') or anxiety felt specifically over political knowledge testing[23]. Through these measures, I am able to perform a first, admittedly approximate, test of mediation via a set of regressions, to assess if (a) treatment impacted anxiety, (b) anxiety impacted confidence about political knowledge, and finally, (c) the effect of treatment is captured entirely by anxiety, when both variables are used simultaneously to predict confidence: a popular approach in mediation analysis [162],[163],[164]. Strong evidence of mediation is an affirmative answer to all the three questions (a)–(c). In this approach, the mediator measure is observed after treatment delivery, thus randomisation inference does not apply on this variable to make a confident causal claim. It might be the case that role model effects impact a third, unmeasured variable which is responsible for both the anxiety and the efficacy effects, thus merely removing anxiety would not result in improved efficacy *per se*.

To strengthen claims of causality, in *Study 2*, a novel 'crossover' experimental approach [165],[164], I introduce experimental manipulation on the mediator variable, anxiety itself, after delivery of the picture treatment. Participants were randomly reassigned into an anxiety group where they were invited to list and reflect on things or situations that make them anxious, often used in experiments to instil state anxiety. A control group listed and elaborated on objects they have recently purchased online[24]. In this framework, men only exposure serves as a control group both in terms of picture treatment and in terms of mediation. Beyond 'men only' exposure, all 'treated' subjects, thus those viewing at least 20% of women or more in picture stimuli, were assigned either in the anxiety group or remained in a control. Without role model effects reversed by anxiety, *only members of the control (online shopping) group* are expected to report higher confidence in political knowledge, their anxieties being successfully alleviated by exposure to role models. By contrast, if viewing gender balanced pictures first, but subsequently members of the anxiety group, role model effects on test subjects are expected to dimin-

ish via re-induced anxiety, and confidence about political knowledge would be reported to comparable levels to those who viewed 'men-only' or low gender-balance pictures. The middle panel of Figure 5.1 is a visual representation of this approach.

Study 1

| Women rep − | → | Anxiety + | → | PE − |

| Women rep + | → | Anxiety − | → | PE + |

Study 2

| Women rep − | → | Anxiety + | → | PE − |

| Women rep + | → | Control | → | PE + |
| | → | Anxiety + | → | PE − |

Study 3

| Anger + | → | AE + |

| Sadness + | → | AE − |

| Control | → | AE − |

[_ _ _] Manipulated [___] Observed/Hypothesized
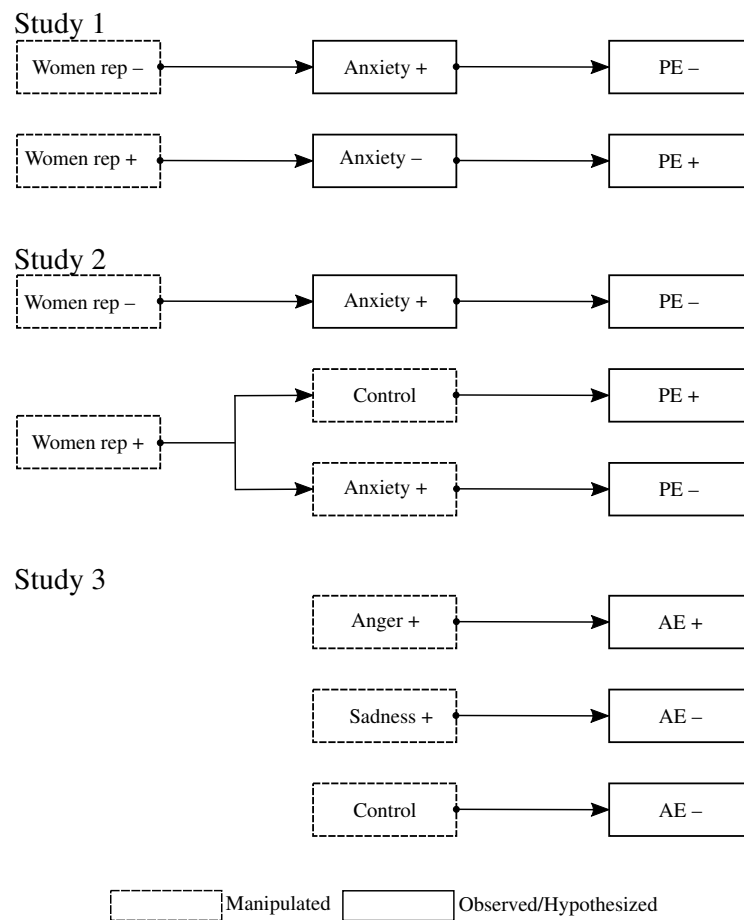
PE *stands for 'political engagement',* AE *for 'attentional engagement'*[25]. $+-$ *signs show hypothesized (if observed) or manipulated direction of effects on women respondents.*

Figure 5.1: Chapter 5—Overview of studies

In *Study 3*, I revisit the eye-tracking study of attentional engagement, to investigate mediation on attentional bias to gender-balance in political groups. Affect remains experimentally manipulated, similarly to Study 2, but with more differentiation in type and a much clearer target. In one group, female participants are invited to reflect on their experience with everyday sexism in particular and on those filling them with *anger*. In another group, they focussed on experiences with sexism resulting in feelings of *helplessness and sadness*. While originally I aimed to keep to the distinction between anger and anxiety, my exploratory study[26] concluded it may be difficult to invite participants to reflect vaguely on 'anxiety' in the context of sexism, but the contrast between action-oriented negative feelings (anger) and feelings where action did not seem possible (sadness, helplessness) were more typical on the topic. I elaborate more on theory regarding emotions of approach and avoidance at the end of Chapter 2 Section 2.1.3.

The following sections report the results of Studies 1–3. To keep to a unified set of statistical tools across my chapters, treatment/control group means per gender (if appli-

cable) are estimated via MCMC using flat priors, however, the classical or frequentist approaches to estimate group means and variances were confirmed to produce identical results. Since these studies are less exploratory in nature than those in Chapter 4 but test specifically for mediation, there is no model expansion from 'null' or uninformative effects.

5.1 Study 1: Self-assessed anxiety as mediator

Anxiety has been central to the theory of stereotype threat since its inception [32]. In the original studies, self-reported anxiety is linked to both receiving the negative treatment (that is, stereotype threat) and to weak performance on standardised testing, while those in the control group not experiencing stereotype threat typically report less anxiety and perform better [35],[37]. Although treatment is defined differently, Study 1 does follow this basic analytical path to demonstrate a link between gender balance in political office and anxiety levels to infer on stereotype threat, and to seek evidence that alleviating anxiety results in greater political engagement, specifically confidence about political knowledge (see section above), thus isolating stereotype threat as mediating mechanism in role model effects.

5.1.1 Data & procedures

In this round of data collection (in previous chapters I called this the second or 'replication round' running through May 2017), I ran an online experiment in which $N = 667$ participants with a near equal gender split were randomly assigned to view and rate a series of pictures with politicians that were either only men, or featuring a token woman, a Critical mass of women, or a Parity of men and women. Immediately following treatment delivery, participants answered a question about state emotion (a self-assessment of emotions felt 'right now') or about testing emotion, latter being an assessment of how they feel about the upcoming political knowledge test. The random split between these emotion questions was to ensure participants do not have to answer two similar questions subsequently, minimising the possibility of projecting responses from one question to another, a strategy they might adopt to provide consistent answers across similar questions[27]. Finally, dependent measures of political engagement were taken including a political knowledge test where 10 UK and international politicians needed to be matched with their jobs, along with an assessment after each question about how confident respondents were in their test answers (in the following referred to as 'confidence about test answer', the dependent variable of this study).

The two emotion questions comprise identical check lists of emotions (respondents ticking the boxes next to each they feel), taken from established inventories of anxiety (STAI, [166]): pleasant, relaxed, calm, uncomfortable, uneasy, worried, nervous, jittery, indecisive, anxious and uncertain. For state anxiety, these were phrased to indicate they felt these emotions 'right now', for testing anxiety, the phrase 'Regarding the test of political knowledge' was added. These together are meant to measure the same underlying

Table 5.1: 'Regarding the test of political knowledge, would you say you are experiencing any these emotions right now?'

Treatment	Gender	% Pleasant	% Relaxed	% Calm	% Uncomfortable	% Uneasy	% Worried	% Nervous	% Jittery	% Indecisive	% Anxious	% Uncertain
Men only	Men	60.00	80.00	85.00	15.00	10.00	20.00	15.00	10.00	20.00	15.00	40.00
Token		72.00	80.00	82.00	20.00	18.00	16.00	14.00	12.00	18.00	14.00	28.00
Cmass		56.52	73.91	73.91	19.57	26.09	23.91	23.91	13.04	28.26	23.91	41.30
Parity		56.00	76.00	76.00	20.00	20.00	26.00	24.00	8.00	30.00	26.00	46.00
Men only	Women	43.48	43.48	43.48	39.13	43.48	43.48	43.48	21.74	47.83	43.48	69.57
Token		51.22	63.41	73.17	24.39	29.27	29.27	24.39	19.51	41.46	29.27	53.66
Cmass		51.11	75.56	77.78	13.33	24.44	20.00	15.56	6.67	17.78	22.22	48.89
Parity		66.67	73.68	71.93	19.30	22.81	19.30	26.32	15.79	29.82	19.30	42.11
N	All	338.00	338.00	338.00	338.00	338.00	338.00	338.00	338.00	338.00	338.00	338.00
$\chi^2(3, N=168)$	Women	5.25	8.07	8.84	6.50	3.90	5.60	6.62	4.00	8.71	5.20	5.12
p		0.15	0.04	0.03	0.08	0.27	0.13	0.08	0.26	0.03	0.16	0.16

Table 5.2: 'Would you say you are experiencing any these emotions right now?'

Treatment	Gender	% Pleasant	% Relaxed	% Calm	% Uncomfortable	% Uneasy	% Worried	% Nervous	% Jittery	% Indecisive	% Anxious	% Uncertain
Men only	Men	77.78	88.89	100.00	11.11	5.56	5.56	0.00	5.56	11.11	0.00	22.22
Token		75.00	88.64	84.09	11.36	15.91	13.64	9.09	9.09	22.73	15.91	25.00
Cmass		72.73	87.27	76.36	10.91	14.55	21.82	9.09	7.27	21.82	18.18	18.18
Parity		64.00	88.00	98.00	6.00	6.00	14.00	8.00	6.00	18.00	14.00	24.00
Men only	Women	83.33	77.78	83.33	5.56	5.56	16.67	11.11	5.56	11.11	22.22	16.67
Token		73.47	75.51	85.71	8.16	12.24	22.45	18.37	4.08	10.20	12.24	26.53
Cmass		66.67	74.51	80.39	25.49	23.53	19.61	19.61	15.69	23.53	17.65	27.45
Parity		67.50	67.50	75.00	17.50	25.00	20.00	17.50	20.00	27.50	25.00	45.00
N	All	329.00	329.00	329.00	329.00	329.00	329.00	329.00	329.00	329.00	329.00	329.00
$\chi^2(3, N=159)$	Women	2.00	0.87	1.56	7.24	5.12	5.60	0.31	0.69	6.48	5.531	5.71
p		0.55	0.83	0.66	0.06	0.16	0.95	0.87	0.09	0.14	0.48	0.12

Note: participants were either exposed to test emotion (Table 5.1) or state emotion (Table 5.2) questions, via random assignment.
χ^2 shows chi-square test statistic for subsample of women, and p the respective p-values.

construct, namely anxiety, tapping into its varying intensities to obtain a continuum as some of these emotions are expected to be less striking and thus easier to subscribe to (e.g. calm) than others (e.g. anxious). Inherently, testing anxiety *is* state anxiety, but a more situation specific version of it evoked by fear of performing badly on a test. For a discussion of situation-specific anxiety approaches in the stereotype threat literature, refer to Chapter 2 Section 2.1.3.

Tables 5.1 and 5.2 on page 95 show the percentage of respondents selecting each particular emotion, demonstrating a number of key patterns. Firstly, there is a clear gender gap in testing anxiety but not in state anxiety. Regarding the political knowledge test, often twice as many women as men respond with anxiety, holding across the majority of anxiety markers. There may be some gaps within the state anxiety markers as well (see 'uncomfortable'), however, these differences do not hold across all anxiety markers and sometimes go the opposite direction (such as 'anxious' against 'uncomfortable'). Secondly, the raw percentages already suggest treatment effects on testing anxiety (tested in sections below) gradually following patterns in gender balance. Women do not only become less anxious across all of the indicators but by parity their anxiety seems to have hit bottom, being comparable to men's initial levels. Thirdly, men's testing anxiety has increased as the proportion of women did in the picture treatment. Fourth, both men and women may have become more anxious with regards to state anxiety; the between-treatment movements being much smaller and not visibly dependent on respondent gender.

To test anxiety as a mediator relying on this set of measures, I follow an analytical strategy most common in regression-based mediation analysis [162], in one instance used to explore mediation of role model effects as well though on different mediating mechanisms [19]. Firstly, I regress Treatment on Effect, namely refit Model 12 in Chapter 4, to find out if increasing gender balance in the picture pool increased women's confidence about their political knowledge. Note that the answer in Chapter 4 is one across both data collection rounds, but here I need to establish the specific treatment effect in the current data set only. Secondly, I take the first half-step to mediation analysis and regress Treatment on Mediator, to find out if increasing numerical balance in pictures decreased (state and testing) anxiety. Thirdly, I take the second half-step in mediation analysis and regress Mediator on Effect, to find out if anxiety decreased confidence about political knowledge. Finally, I run an analysis of the full, mediated path in which both Treatment and Mediators are regressed on the Effect, where suggestive of mediation is if the gender balance effect disappears via inclusion of self-assessed anxiety, as a result of the mediator variable capturing enough of the treatment effect.
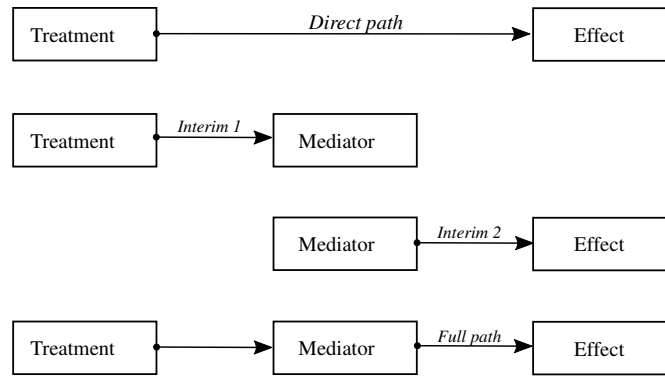
Treatment is increased gender balance in picture stimuli, mediators are state anxiety and testing anxiety. Effect is 'Confidence about political knowledge'.

Figure 5.2: Chapter 5 Study 1—Steps of mediation analysis

Table 5.3: Descriptive statistics: Respondent N & Mean Conf. in political know.

Treatment	Gender	Respondent N	Mean confidence
			(across 10 scales each ranges 1–10)
Men only	Men	38	7.13
Token		36	7.83
Critical mass		35	7.33
Parity		31	6.04
Men only	Women	41	5.72
Token		31	5.68
Critical mass		28	6.08
Parity		36	6.76

5.1.2 Results

Step 1, Direct path: From cause to effect

Chapter 4 found most evidence that an increasing proportion of women in the picture pool increased women participants' confidence in their political knowledge, measured across confidence ratings following each test question. To that end, it used a combined sample generated from an 'original round' of data collection and a 'replication round' of data collection. Specifically this meant pooled analysis of 605 subjects' self-assessed confidence recorded across 10 political knowledge test questions, resulting in a final $N = 6,708$ sample size in the item-level or 'stacked' data set for analysis.

Here I refit this model using the the 'replication round' sample, without pooling, where the emotions questions were asked to investigate mediation. For this model, subject N drops to 276, with the item-level or stacked N being 2,760. Note that the data collection round itself included a larger respondent $N = 667$: currently using the control group of people receiving no further experimental manipulation such as those reported in Chapter 5 Study 2 and Chapter 6 Study 2. Further details about the sample structure are given in Chapter 3, Methodology.

Confidence about test answer
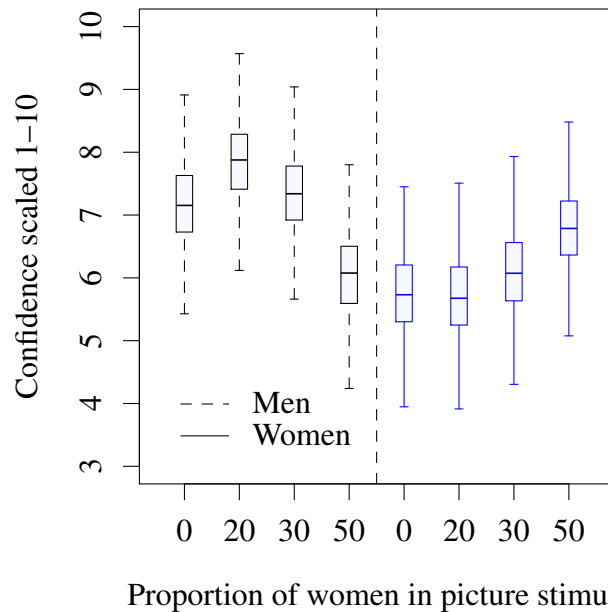
95% distribution of posterior predictions

Proportion of women in picture stimuli

Key: β coefficients (posterior means) and 95% CI corresponding to the effect of women's presence in picture stimuli at the discrete proportions of near 20% or token, near 30% or critical mass and 50% or parity.

Figure 5.3: 'Direct path' results, Model 18

In a 4 treatment groups · 2 gender categories design, after removal of missing data on gender (1 record), cell sizes average at 34.5, determined via randomisation and thus range from 28 (women in the Critical mass group) to 41 (women in the men only group). Confidence is indicated on 10 point scales. In Chapter 4, I used standardised scores to be able to compare effects across several dependent variables. In this chapter, my DV is fixed across Studies 1 and 2, therefore I will use the untransformed raw scores in a linear, multilevel regression fit by MCMC to derive the 8 group means and infer on treatment effects.

Model 18 fits this model, results are reported in the first column of Table 5.5 on page 5.5. Replicating the findings in the pooled sample in Chapter 4, I find strong evidence that though women are less confident than men are in their test answers if viewing men only pictures (estimated at -1.42 points on a 10 point scale), the confidence gap disappears gradually as gender balance in picture stimuli increases and women catch up to, if not surpass, men's confidence levels by parity (see also 'Models 18–21: Summary of key effects' on page 104). Figure 5.3 shows the treatment group-level predictions using the 95% posterior distribution, showing relatively wide prediction intervals due to large unexplained variance, nevertheless demonstrating treatment effects.

Step 2, Interim link: From treatment to mediator

In this section I explore treatment effects on anxiety, an essential link to establish if the absence or low proportions of women in visible political office may translate into stereotype

threat, and to see if improving gender balance successfully alleviates this threat. On the one hand, I keep the split between testing-specific anxiety and more diffuse state anxiety (see Tables 5.1 and 5.2) by analysing them separately; on the other, I pool the 11 anxiety markers together within these two, treating them all as tapping into the same underlying construct, namely either 'state anxiety' or 'test anxiety'.

I begin with a confirmatory scaling exercise testing this unidimensionality via Mokken scaling [167],[168, for R package]. I obtain high scale scalability coefficients [169] $H = 0.57(SE = 0.03)$ for state anxiety and $H = 0.62(SE = 0.03)$ for testing anxiety suggesting good measurement validity. Pairs of items, if measuring similar latent constructs, should and do co-vary highly, across both anxiety types only 10 out of 110 possible pairings having normed covariance less than 0.50.

Given satisfactory scale performance, I am confident in pooling together the single item responses into a stacked data set similar to the one used in the previous section, and thus obtain a sample structure where 11 emotion responses are nested within $N = 327$ individuals if testing treatment effects on state anxiety, and $N = 336$ individuals if testing treatment effects on testing anxiety, resulting in stacked N's for analysis of 3,597 and 3,696, respectively. Respondent N's increase compared to the 'direct path' model: while there the dependent measure, confidence in test answers, was taken after further experimental manipulation, irrelevant for the direct path model and thus excluded from the sample, state anxiety *or* test anxiety measures were taken before the delivery of those further manipulations thus all subjects remain in this sample.

Using this stacked data set, Models 19.1 (state anxiety) and 19.2 (testing anxiety) are multilevel logistic regression models fitted with MCMC to derive the probability of an anxious response within 8 treatment x gender groups (fixed effects) and controlling for unequal variances on the individual and test item levels (random effects). I expand on the full results in the first two columns of Table 5.4. Broadly, treatment effects on state anxiety do not hold, while treatment effects on Testing anxiety do stand out. In other words, there is no evidence that either of the treatment groups instilled or decreased self-assessed state anxiety but there are gradual effects on testing anxiety. The direction of these effects depends heavily on respondent gender. Men become *more* anxious about political knowledge testing if viewing more women in picture stimuli. The average treatment effect on men under parity exposure is over one and a half times ($e^{0.47} = 1.60$) more likely anxiety than under men only picture exposure, though still sizeable posterior probability (about 15%) of null or negative effects. Effects are much more robust on women, the odds of an anxious response decreasing to about a tenth of that observed under men only exposure. The predicted proportion of women, as visualised by Figure 5.4, decreases from 45% to 17% by token, 10% by critical mass, and 9% by parity. There is a small residual variance left in terms of odds in Model 19.2 suggesting good model fit, especially contrasted with that of the state anxiety model.

This concludes the test of emotion as a dependent variable, to be integrated into the full mediation model in Step 4. Though no treatment effects were detected on self-

Testing anxiety

95% distribution of posterior predictions
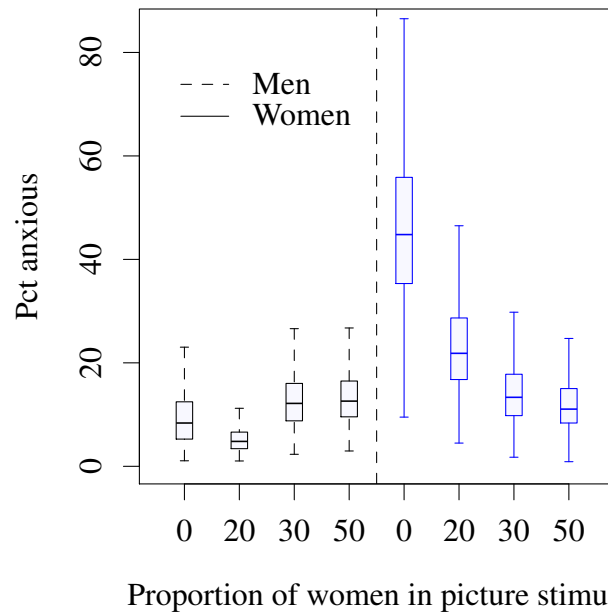
Key: *β coefficients (posterior means) and 95% CI corresponding to the effect of women's presence in picture stimuli at the discrete proportions of near 20% or token, near 30% or critical mass, and 50% or parity.*

Figure 5.4: 'Interim link 1' results, Model 19.2

assessed state anxiety, for full information, the latent variable will be included along with testing anxiety in the further models even though current results cast doubt on its constituent role in mediating role model effects.

Step 3, Interim link: From mediator to effect

In this model, anxiety is the independent and confidence is the dependent variable. Similarly to the direct path model, this model will make use of a narrower subject pool $N = 276$ (see above) but since state or testing anxiety were presented as alternative questions, the former model will use subject $N = 130$ and latter $N = 146$, resulting in stacked N's across confidence ratings through 10 test questions of 1,300 and 1,460, respectively.

In this setup, however, anxiety needs to be a single independent latent variable on the individual level as the more efficient stacking approach is only feasible if applied to latent *dependent* variables. To this end, data reduction techniques such as factor analysis are available. Given discreet data observed through dichotomous response patterns (respondents either 'felt' given emotion or 'did not feel' given emotion, I use an Item Response Theoretical framework to generate this latent variable given the 11 individual emotion-item responses [170],[171, for R package].

Latent trait models under IRT such as the current ones for testing- and state anxiety are widely used in psychometrics research specifically to evaluate scale validity, an issue I addressed briefly in the previous section via reporting Mokken's scalability coefficients. These models estimate further useful information, not of immediate interest in this sec-

tion, such as variation across single items, in my case emotion markers, due to item 'difficulties' (how typically or rarely 'felt' an emotion is in the sample) as well as due to items' 'discriminatory power' (how well a particular item distinguishes between levels of anxiety, for example, is the particular item needed to describe extreme levels of anxiety or alternatively, is it simply needed to split between high and low anxieties). These IRT models find that that there is relatively little variation across coefficients associated with item difficulties for state anxiety (0.91–1.52), but somewhat more in testing anxiety (0.16–1.25), the extremes being 'uncertain'—a typical choice among respondents— and 'jittery'— choice not very typical among respondents. Similarly, discrimination parameters vary more in testing anxiety than in state anxiety in which 'indecisive' ($\beta = 1.35, SD = 0.20$) may not discriminate extreme high anxiety individuals from moderate anxiety individuals very well, as opposed to the reverse coded 'relaxed' ($\beta = 6.70, SD = 1.97$).

Via this model, I am able to derive, for each respondent, a factor score to describe his or her level or intensity of anxiety. These scores are estimates themselves, subject to standard error but are nevertheless useful summary estimates of the level of anxiety felt by test subjects. These scores are standardised, those around 0 indicating closeness to an estimated population-level average, and negative values indicating below-average, positive values indicating above-average state or test anxiety. State anxiety factor scores range from -0.54–1.73 with a mean of -0.07 and a standard deviation of 0.59, test anxiety factor scores range from -0.80–1.68, with a mean 0.14 and a standard deviation of 0.79.

These scores are thus imputed in the stacked data set to fit Model 20.1 with state anxiety and Model 20.2 with testing anxiety, both being multilevel linear models using the 1–10 test answer confidence ratings as dependent variable, fit by MCMC, random effects defined as test questions (confidence submitted across 10 questions in the political knowledge quiz) and subjects to reflect nesting structure. Full results are reported in Table 5.4.

Model 20.1 using state anxiety as a predictor, does not find enough evidence of an effect on confidence, the posterior mode of the of the 'Anx state' Beta is essentially zero, surrounded by large credible intervals. Model 20.2, however, finds a testing anxiety effect predicting a decrease in confidence of 0.74 (confidence measured on a 10 point scale) corresponding to a unit increase in latent anxiety. Using the empirical range of the anxiety scores (see paragraph above), respondents' predicted confidence at low anxiety (-0.80) is around 7.55 (well above average) whereas for a high anxiety individual it is predicted at 5.77 (below the average of 6.33 across both genders). Figure 5.5 plots this regression line. Interestingly, a closer inspection of random effects show that anxiety over the political knowledge test left still a fair portion of residual variation in test answer confidence.

Step 4, Full mediated path

The previous models presented necessary evidence that women's self-assessed anxiety, expressed specifically over political knowledge testing, may be a mediator of stereotype threat effects and results in uncertainty about the test answers themselves. Contexts where

Confidence about test answer

95% distribution of posterior predictions

Anxiety low–high (latent trait)
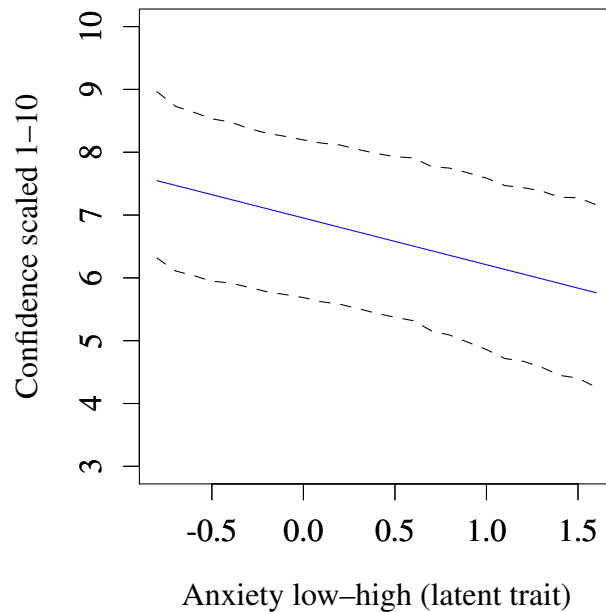
Key: IV is latent (z) scores, positive values indicating greater-than-average testing anxiety.

Figure 5.5: 'Interim link 2' results, Model 20.2

role model effects were hypothesized, namely exposure to more women in picture stimuli, translated into less anxiety and more confidence about test answers. These models cast doubt that my self-assessed state anxiety measure works as a mediator.

Models 21.1 and 21.2 (full results in Table 5.5) present the fully mediated path right next to the direct path, presenting a regression in which both treatment and mediators, state anxiety and testing anxiety, are regressed on test answer confidence. Key independent variables are the 8 groups corresponding to combinations of gender x treatment, as well as the additive anxiety effects. As asserted in the regression-based mediation analysis literature, a successful mediator captures treatment effect in its entirety—I will pay attention to both effect sizes and credible intervals—while a partial mediator may capture some of the treatment effect but may also allow the original treatment variable to retain some explanatory power. Thus the basic expectation, if mediation is present, is that gender balance in pictures would *cease to be a predictor* of test answer confidence while (testing) anxiety predicts test answer confidence on its own.

The model-based results confirm this expectation. Although state anxiety does nothing to remove treatment effects, remaining an essentially null effect, the testing anxiety coefficient remains robust and consistently negative within a 95% highest posterior density region. In this model, both gender and treatment lose their magnitude as predictors with credible intervals enclosing zero in all treatment groups. The single-standing gender coefficient is cut by more than a half compared with the direct path model. Though still not exactly zero, it has a much larger posterior variance, suggesting that testing anxiety explains the gender gap as observed especially under men only exposure. A similar pat-

Table 5.4: Models 19.1–19.2 and 20.1–20.2, 'Interim links'

beta	m19.1	m19.2	m20.1	m20.2
DV	Pr(*state anxiety*)	Pr(*testing anxiety*)	*Confidence in test answer, scaled 1–10*	
Type	Binomial / MCMC	Binomial / MCMC	Linear / MCMC	Linear / MCMC
Stacked *N*	3597	3696	1300	1460
Subject *N*	327	336	130	146
Item *N*	11	11	10	10
DIC	1857.16	2453.47	6161.76	6784.76
	Beta [95% CI]	Beta [95% CI]	Beta [95% CI]	Beta [95% CI]
Intercept	-5.45[-7.98;-2.74]	-2.41[-3.84;-1.12]	6.35[5.25;7.60]	6.95[5.69;8.20]
Gender: Woman	0.34[-2.76;3.41]	2.22[0.55;3.83]		
Treatment: 20%	0.57[-2.09;3.05]	-0.58[-1.86;1.00]		
Treatment: 30%	1.21[-1.30;3.72]	0.42[-1.10;1.74]		
Treatment: 50%	0.80[-1.70;3.08]	0.47[-0.49;1.81]		
Woman x 20%	0.43[-3.13;3.86]	-0.49[-2.50;1.47]		
Woman x 30%	0.41[-3.32;3.70]	-2.11[-4.14;-0.21]		
Woman x 50%	1.28[-2.09;4.78]	-2.36[-4.62;-0.56]		
Anx: state			-0.07[-0.72;0.56]	
Anx: testing				-0.74[-1.14;-0.32]
	Gamma [95% CI]	Gamma [95% CI]	Gamma [95% CI]	Gamma [95% CI]
Subject var	14.27[7.50;20.94]	6.80[5.16;8.53]	4.55[3.47;5.94]	3.46[2.52;4.31]
Item var	1.21[0.26;2.75]	0.88[0.22;1.85]	3.11[0.87;6.92]	3.99[1.17;8.76]
Residual var	2.83[0.40;5.50]	0.04[0.02;0.08]	6.09[5.63;6.57]	5.58[5.18;6.03]

tern is observed looking at the treatment coefficients, notably a parity effect halved, large credible intervals not ruling out the possibility of null effects. Comparing the direct path model with the fully mediated path model using testing anxiety, my results thus suggest that testing anxiety mediated role model effects on women's confidence.

Table 5.5: Models 18 & 21.1–21.2, Direct and mediated ('full') paths

	m18	m21.1	m21.2
DV	*Confidence in test answer, scaled 1–10*		
Type	Linear / MCMC	Linear / MCMC	Linear / MCMC
Stacked *N*	2760	1300	1460
Subject *N*	276	130	146
Item *N*	10	10	10
DIC	12958.73	6160.14	6785.28
	Beta [95% CI]	Beta [95% CI]	Beta [95% CI]
Intercept	7.17[5.93;8.68]	6.76[5.25;8.05]	7.34[6.02;8.89]
Gender: Woman	-1.42[-2.27;-0.48]	-1.83[-3.17;-0.49]	-0.72[-1.89;0.58]
Anx: state		-0.17[-0.76;0.46]	
Anx: testing			-0.63[-1.04;-0.15]
Treatment: 20%	0.68[-0.21;1.61]	1.45[0.06;2.86]	-0.07[-1.50;1.03]
Treatment: 30%	0.18[-0.73;1.10]	0.42[-1.07;1.82]	0.30[-0.94;1.52]
Treatment: 50%	-1.13[-2.08;-0.24]	-0.96[-2.15;0.58]	-1.02[-2.28;0.32]
Woman x 20%	-0.74[-2.04;0.60]	-0.84[-2.76;1.09]	-0.33[-2.34;1.33]
Woman x 30%	0.17[-1.16;1.57]	0.54[-1.61;2.69]	-0.58[-2.69;1.08]
Woman x 50%	2.17[0.88;3.40]	2.65[0.66;4.47]	1.35[-0.64;3.06]
	Gamma [95% CI]	Gamma [95% CI]	Gamma [95% CI]
Subject var	3.68[2.97;4.45]	3.83[2.75;5.10]	3.47[2.45;4.37]
Item var	3.47[0.99;7.17]	3.01[0.87;6.59]	3.79[1.12;8.19]
Residual var	5.86[5.53;6.20]	6.08[5.62;6.59]	5.57[5.13;5.94]

Table 5.6: Models 18–21: Summary of key effects

	Model	DV.	Key IV.	β [95% CI]	subject N	stacked N
Direct path	18	Confidence	Parity x Women	2.17 [0.88;3.40]	276	2,760
Interim link 1	19.1	State anxiety	Parity x Women	1.28 [-2.09;4.78]	327	3,597
	19.2	Testing anxiety	Parity x Women	-2.36 [-4.62;-0.56]	336	3,696
Interim link 2	20.1	Confidence	State anxiety	-0.07 [-0.72;0.56]	130	1,300
	20.2	Confidence	Testing anxiety	-0.74 [-1.14;-0.32]	146	1,460
Full path/ state a.	21.1	Confidence	Parity x Women	2.65 [0.66;4.74]	130	1,300
	21.1	Confidence	State anxiety	-0.74 [-0.76;0.46]	130	1,300
Full path/ test a.	21.2	Confidence	Parity x Women	1.33 [-0.38;3.10]	146	1,460
	21.2	Confidence	Testing anxiety	-0.63 [-1.04;-0.15]	146	1,460

Full results at respective model tables. '95% CI' is the 95% highest posterior density area of the predictors. 'Stacked N' is the number of test items or emotion items x number of respondents, after missing data exclusion.

5.1.3 Section Conclusions

In this study, I presented results of a regression-based approach to mediation analysis broken down into four steps. In Step 1, I confirmed that my treatment, an experimentally improved gender balance in picture stimuli depicting political campaign groups, impacted if respondents gave a confident answer to their political knowledge test questions. As suggested by the role model hypothesis, women were more confident across 10 test answers if looking at political groups featuring a higher proportion of women. At the same time, men gave less confident answers although, looking at the model-based predictions, they may have experienced a confidence boost when exposed to a token female presence.

In Step 2, I looked at self-assessed anxiety expressed through 11 emotion markers measuring either diffuse state anxiety ('Would you say you are experiencing this emotion now?') or testing anxiety ('Regarding the upcoming political knowledge test, would you say you are experiencing this emotion now?') and confirmed that treatment effects impacted the assessment of testing anxiety: women gave systematically less anxious responses on these emotion markers if viewing pictures with improved gender balance previously.

In Step 3, I regressed both these types of anxieties on confidence and found that testing anxiety, not state anxiety, predicted test answer confidence. A potential concern is whether 'testing anxiety' and 'test answer confidence' may capture too similar concepts that the relationship between them is perhaps definitional, rather than causal. My results in terms of residual variance, however, suggest that a variance amounting to almost half of the full range of the dependent variable (5.58, or a standard deviation of 2.36) is left unexplained after factoring in testing anxiety.

Finally, in a fully mediated path, my model presented evidence that gender and treatment effects on confidence may be captured by testing anxiety, establishing this emotion as a mediator of treatment effects. A summary of the key effects within these paths is displayed in Table 5.6.

Drawing heavily on the stereotype threat and role model literature, I treat this as pre-

liminary evidence that political role models, potentially women in visible political office, can alleviate stereotype threat effects in the political domain, such as women's potential disengagement from politics. That stereotype threat is a relevant concept in political behaviour, is captured by pointing at anxiety itself as mediator: according to this interpretation, political engagement may be driven not only by policy considerations or even partisanship but also by more automatic, affective processes.

Before drawing broader conclusions about the role of stereotype threat in women's political engagement, however, there are potential limitations to point to. First, although this study utilised experimental manipulation of treatment in a survey experiment setup, it still relies on observational data about emotion. While the process of mediation seems plausible looking at the regression results, and presents sufficient evidence to the causal 'primacy' of treatment (proportion of women politicians), it leaves the relationship between mediator, effect and unobserved variables, unaddressed. It remains open whether treatment impacted some other variable that manifests both in anxiety and confidence.

Additionally, self-assessed emotions are subject to misinterpretation and misreporting on the respondent's behalf. For instance, testing anxiety inherently *is* state anxiety in that it is momentarily felt anxiety. Differences emerge in that the state emotion was diffuse, not mentioning any target, but testing anxiety was an explicit reminder of the upcoming test. When at the point of testing, those reporting no state anxiety have likely generated some anxiety. Looking at emotion directly at testing without relying on self-report is however not feasible in an online experiment.

To overcome at least some limitations in the observational approach, and to further examine the role of state anxiety in its more general sense, in Study 2 I undertake a different approach in mediation analysis. Via experimental manipulation of state anxiety before testing, I am seeking to introduce a more appropriate test of causality. Though limitations still apply, especially at the point that more reliable—such as psycho-physiological—measures of anxiety are not taken in an online survey setup, I will be able to use Study 2 as robustness check of results in Study 1 and potentially clear up if differences between state anxiety and testing anxiety were purely methodological or more substantive.

5.2 Study 2: Experimental anxiety as mediator

Most studies test underlying mechanisms by observation. In randomised trials, following treatment delivery, it is very straightforward to collect measures not only on a hypothesized effect but on a hypothesized mediator as well and compare both across treatment groups to infer on mediation. In Study 1, this meant looking at self-assessed anxiety, measured after delivery of picture stimuli, and I concluded they lent support to the hypothesis that role models increased women's confidence about their political knowledge via successfully alleviating anxieties over political knowledge testing.

There were, however, a number of limitations I identified. First, while the causal mechanism seems, on the surface, tenable because of treatment randomization, *there is no comparable randomization inference applicable to anxiety*; anxiety levels were self-

reported but not experimentally manipulated. Without an element of randomization on anxiety alternative and unobserved mediators affecting both emotion and the dependent variable cannot be ruled out.

Second, though my analysis found 'testing anxiety' to mediate stereotype effects, it failed to uncover effects through more diffuse 'state anxiety' self-reports which would lend further support to the anxiety hypothesis altogether. I concluded that differences between testing and state anxiety may lie in one of them including a reminder of the upcoming political knowledge test, which may have 'helped' self-report in giving an assessment of anxiety, while the other did not include anything to assess anxiety about, which may have prevented accurate appraisal of emotion. Nevertheless, further tests on anxiety are needed to clarify why the discrepancy between these two measures emerged.

Third, I pointed out that in the web survey context I could only collect self-report measures about anxiety, though across multiple emotion markers. There are a number of obvious limitations relying on these variables, and in the stereotype threat literature these issues have repeatedly emerged when self-report led to inconsistent or null results while physiological measures often revealed heightened anxiety [34].

Though statistical adjustments are available to reduce bias stemming from failure to include unobserved variables, [172],[173] in Study 2 I overcame the first problem via research design, rather than *post hoc* adjustments, and introduced randomisation on the mediator variable itself. Essentially, treated subjects—viewers of role model pictures with at least a token woman in them—are randomly reassigned into a control and an anxiety group carrying out two writing tasks that were meant to induce state anxiety. By reflecting on anxious experiences, I obtain a re-induced anxiety group where role model effects may vanish, if the anxiety mediation hypothesis stands.

The second and third problems are addressed partially via a different measurement strategy where self-report is not used. Strict separation between state anxiety and testing anxiety also disappears and the focus in the writing task is on state anxiety in general. The next sections review this design and the results. Table **??** below shows possible combinations of treatment and mediation groups. Within the framework of Chapter 5, see also 'Chapter 5—Overview of studies' on page 93 for a better comparison with the rest of the studies.

Table 5.7: Chapter 5 Study 2—Experimental design

| | | Picture treatment | |
		Control *'Men only'*	Treatment $\geq$ *'Token'*
Mediation	Control	Confidence −	Confidence +
	Anxiety	x	Confidence −

Treatment is exposure to pictures of groups of politicians featuring 0% men only, 20% token, 30% critical mass, or 50% parity presence of women. Randomised mediation induces state anxiety: anxiety group carried out two writing tasks reflecting on 'things that make [them] anxious', control group on items recently bought online. Dependent variable is confidence about political knowledge, +/- signal hypothesized direction of effects on women.

5.2.1 Data & procedures

This study uses survey experiment data from the 'replication round' of data collection where mediation was also investigated, overlapping with the previous study with one important distinction: here I am able to use a larger sample of $N = 471$ respondents, keeping a near equal gender split. By contrast, by the end of Study 1 sample sizes for both state anxiety models and testing anxiety models were below 150 as those numbers were exclusive of the number of participants analysed here in the 'anxiety group', see below and Chapter 3 Figure 3.4 on page 3.4 depicting sample management.

Table 5.8: Descriptive statistics: Respondent N & Mean Conf. in political know.

Treatment	Mediation	Gender	Respondent N	Mean confidence (1–10)	St.Dev.
Men only	(Control)	Men	38	7.13	3.47
Token	Control		36	7.83	3.06
Token	Anxiety		34	7.43	3.25
Critical mass	Control		35	7.33	3.22
Critical mass	Anxiety		33	7.39	3.34
Parity	Control		31	6.04	3.64
Parity	Anxiety		32	7.32	3.26
Men only	(Control)	Women	41	5.72	3.65
Token	Control		31	5.68	3.60
Token	Anxiety		31	6.37	3.63
Critical mass	Control		28	6.08	3.58
Critical mass	Anxiety		33	5.94	3.77
Parity	Control		36	6.76	3.28
Parity	Anxiety		32	6.89	3.27

After picture exposure, participants were randomly reassigned into a control and an anxiety group[28]. All of them had to carry out two writing tasks: in the control group, participants were invited to first list a few things they had recently purchased online, and then were asked to describe one of those items more in detail. In the anxiety group, participants were invited to list a few things that make them anxious, and were then asked to elaborate on one of the things/experiences listed. This basic writing task was meant to induce state anxiety, a method to create experimental emotion with relatively consistent results or 'compliance' rates in previous studies [174].

An important detail is that, due to sample size limitations, participants in the 'men only' picture-viewing groups did not get sorted into an anxiety group. Besides operational reasons, this is also defensible on theoretical grounds: via mediation I test how and why treatment works, rather than the absence of the treatment. The hypothesis about mediation is that role models *alleviate* anxieties, however, there is no explicit hypothesis about mediation in the de facto 'control' or untreated group, which here is viewers of 'men only' pictures. It is this group where highest anxiety levels by women are hypothesized (and also observed in Study 1), and treatment is concerned with the removal of this negative emotion. Rather than inducing anxiety, it would have made sense to *remove* anxiety experimentally in this group, which remains outside of the scope of this study.

Experimentally created anxiety is hypothesized to re-induce anxiety in the treated (viewers of role model pictures) and thus diminish role model effects on test answer confidence. The group not receiving experimental anxiety induction would then perform as expected in Chapter 5 Study 1 and Chapter 4: women viewing role model pictures would give more confident test answers, while those viewing men only pictures would give less confident answers.

Table 5.8 is an overview of test answer confidence descriptively per combinations of treatment x mediation categories. These averages suggest that maybe mediation effects do not follow the hypothesized path: looking at women, though control group averages follow the role model effect (peaking at parity), the anxiety group averages show that perhaps (a) 'anxious' women were more confident under token woman presence than in the control group and that (b) 'anxious' women may be comparably confident to their control counterparts under parity presence in the picture stimuli. While options (a) and (b) both suggest that role model effects prevail *despite* re-induced anxiety, these averages across all confidence ratings may hide important variations and thus I will look at the model-based answers to determine if state anxiety stands as a mediator.

The basic analytical strategy to follow may be simpler than the strategy introduced in the previous, regression-based mediation study in that it is feasible in single step of estimation to derive 2 x 7 group means: for each gender, one 0% or men only group mean (the baseline, control group), and for 20% or token, 30% or critical mass, and 50% or parity treatment groups, a group mean under control (online shopping) and one under anxiety. To be able to make the appropriate comparisons of treatment and anxiety effects within gender (and foremost women), however, I am presenting results step-by-step in smaller and simpler models, namely

1. Model 22 estimating the effect of anxiety on all treated women *across treatment groups*, where baseline treatment category (intercept) is 20% token;

2. Model 23 estimating the effect of anxiety on all treated women *within treatment groups*, , where baseline treatment category (intercept) is 20% token;

3. Model 24.1 predicting *all* (treated as well as untreated) 7 group averages for women respondents, , where baseline treatment category (intercept) is 0% men only;

4. Model 24.2 predicting *all* (treated as well as untreated) 7 group averages for men respondents, , where baseline treatment category (intercept) is 0% men only.

All key quantities above are estimated via MCMC in multilevel models where 10 confidence ratings are nested within respondents and test questions to account for dependence among observations within individuals and within particular questions. Note that the dependent variable *cannot* be estimated via three-way interaction between treatment, gender, and mediator: Such model would be 'rank deficient' because not all possible combinations of these three grouping factors are populated as men only control group respondents never got to be assigned into the anxiety group but only in the control group.

5.2.2 Results

Model effects are interpretable on the original confidence scale with a range of 1–10. Table 5.9 on 111 lists the full results of all models fitted in this section. The baseline category of the multi-category treatment variable varies: the model looking at mediation effects only on the treated subsample of women uses 20% token as baseline category thus Beta 'Intercept' in Models 22 and 23 shows average confidence under token presence in the control (online shopping) group. In subsequent models using both treated and untreated samples of men and women, this baseline category remains the men only picture viewing group.

Overall, my results suggest that the anxiety manipulation did not diminish role model effects, contrary to theoretical expectations, but consistent with Study 1 findings about state anxiety. All anxiety group predictions about test answer confidence average higher than that of the men only group (the stereotype threat context) and the differences between anxiety and control within each treatment category are negligible, at least for women. This section elaborates on these findings, the next section extends analysis for potential problems in random group assignment, and the last section draws conclusions.

Confidence about test answer

95% distribution of posterior predictions
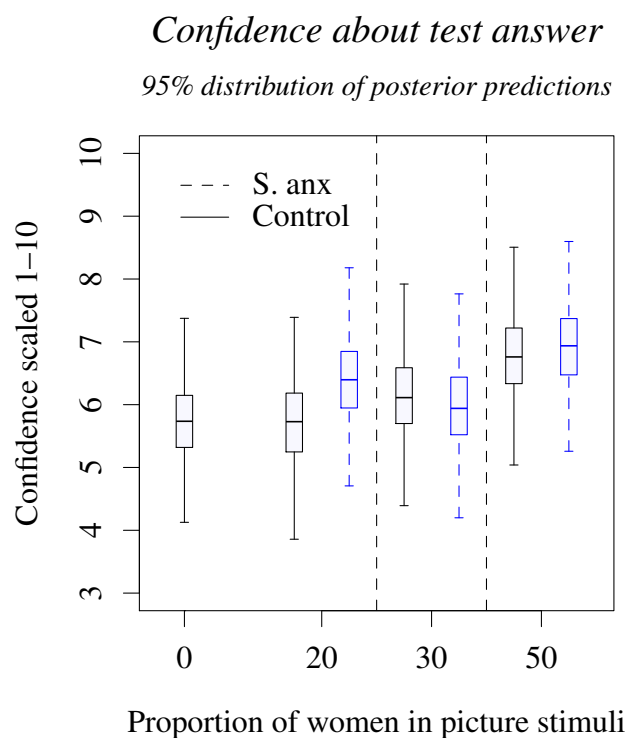
Key: β coefficients (posterior means) and 95% CI corresponding to the effect of women in picture stimuli at the discrete proportions of near 20% or token, near 30% or critical mass, and 50% or parity. Women subsample only.

Figure 5.6: Randomised mediation study, Model 24.1

Anxiety effect on the treated.

Women's mean confidence, *not* accounting for nesting, across all treatment groups not receiving an anxiety manipulation is 6.21 ($SD = 3.50$) as opposed to 6.39 ($SD = 3.58$) if receiving anxiety manipulation, suggesting either null effects or that women were very

slightly more confident under anxiety, pointing in a direction opposite to the hypothesis. An exploration of the number of responses within each confidence level reveals that anxiety if anything *polarised* responses, while in the control group 90 respondents (10% of total) selected the middle point of the confidence scale, only 54 did so in the anxiety group and most of that differential went either in the lowest confidence category (+14 people) or the highest (+30 people). No comparable movements are detectable within men in terms of polarisation but their mean confidence level also increased somewhat by anxiety ($M = 7.11$ and $M = 7.38$, respectively, $SD = 3.38$ and $SD = 3.28$).

Model 22 teases out the anxiety effect on women in a multilevel model accounting for individual and test-question level nesting and finds a near zero effect within wide credible intervals (-0.45 to 0.82) suggesting that the manipulation had no effect on confidence.

Anxiety effects within treatment groups.

Model 23 elaborates on the anxiety effect allowing that it differs depending on the treatment group. Table 5.8 shows mean confidence levels where anxiety is associated with somewhat lower confidence only in the critical mass group for women, further suggesting that any differences between treatment and mediation groups are in fact randomly distributed. The pattern of polarisation on the confidence scale is mostly driven by movements in the critical mass group as well, where about 50% less women selected the middle point of the scale, but two-thirds more selected the lowest point.

The model fails to detect anxiety effects within treatment groups where women are predicted to be comparably confident to the respective control groups. The anxiety x treatment interaction terms are negative, consistent with the hypothesized direction (though leaving predicted anxiety higher within token) but falling within very large credible intervals. Residual variance is estimated exactly at the same value as in Model 22 suggesting no improvements in model fit.

Anxiety effects benchmarked against the untreated.

Finally, Models 24.1 and 24.2 fit the fully specified models using the subsamples of women and men, respectively, where the group means include those under men only picture exposure as well. Presentation of model output changes in the table: while everywhere else, categorical treatment and categorical gender (and categorical anxiety) variables are 'interacted' in the model formula resulting in a prediction at each possible combination, here such categories are combined by hand because some analytically possible combinations are missing, namely, there is no anxiety x men only category. Therefore, in these models the independent variable is a 7-category variable benchmarked to the control (online shopping) x men only picture treatment group.

Figure 5.6 shows the predictions based on Model 24.1 (women) and reveals that the anxiety group predictions closely follow the control group predictions where confidence gradually increases women's confidence by Parity with a marginal effect of $\beta = 1.02$ in the control group and $\beta = 1.18$ in the anxiety group both within positive credible

intervals suggesting consistently positive effects. Because a 7-category variable as solo independent variable has very little hope to predict any dependent measure well, residual variance actually increases compared with the previous models.

Model 24.2 on men's confidence ratings shows that these same group averages are more sensitive to anxiety than women's. While control group averages mirror the findings in Study 1, in that men's confidence *decreased* as a function of treatment, it seems that if anything men *remained confident* if state anxiety was introduced, a rather atypical pattern of performance under anxiety.

Table 5.9: Models 22–23 and 24.1–24.2, Randomised mediation study

beta	m22	m23	m24.1	m24.2
	treated WOMEN	treated WOMEN	all WOMEN	all MEN
DV		*Confidence in test answer, scaled 1–10*		
Type	Linear / MCMC	Linear / MCMC	Linear / MCMC	Linear / MCMC
Stacked *N*	1910	1910	2320	2390
Subject *N*	191	191	232	239
Item *N*	10	10	10	10
DIC	9005.92	9006.55	11007.2	10952.94
	Beta [95% CI]	Beta [95% CI]	Beta [95% CI]	Beta [95% CI]
Intercept	6.21[5.13;7.23]	5.68[4.43;7.12]	5.76[4.59;7.15]	7.15[5.83;8.49]
Mediator: s. anxiety	0.18[-0.45;0.82]	0.69[-0.39;1.85]		
Treatment: 30%		0.40[-0.82;1.60]		
Treatment: 50%		1.09[0.06;2.30]		
Anx x 30%		-0.80[-2.38;0.78]		
Anx x 50%		-0.55[-2.01;1.14]		
Group 20 & Control			-0.02[-0.98;1.13]	0.71[-0.19;1.65]
Group 20 & Anx			0.66[-0.42;1.62]	0.29[-0.70;1.13]
Group 30 & Control			0.37[-0.52;1.50]	0.21[-0.71;1.02]
Group 30 & Anx			0.21[-0.78;1.17]	0.25[-0.67;1.14]
Group 50 & Control			1.02[0.14;2.01]	-1.08[-2.04;-0.15]
Group 50 & Anx			1.18[0.18;2.22]	0.19[-0.68;1.09]
	Gamma [95% CI]	Gamma [95% CI]	Gamma [95% CI]	Gamma [95% CI]
Subject var	4.62[3.52;5.70]	4.57[3.55;5.73]	4.26[3.44;5.17]	3.41[2.69;4.14]
Item var	2.97[0.68;6.24]	3.11[0.91;6.62]	3.36[0.79;7.40]	3.60[0.85;7.08]
Residual var	5.96[5.57;6.37]	5.96[5.62;6.37]	6.14[5.78;6.49]	5.24[4.95;5.54]

5.2.3 Measurement validity: Emotion through self-report and text

Given anxiety's null effects on women, and its atypical effects on men, questions arise as to the success of the randomised writing task, namely about success in randomisation, about success in creating experimental anxiety, and about respondents' motivation to control behaviour.

To address the first concern, I am able to use self-reported test anxiety measures, taken *before* randomised anxiety induction, to predict anxiety or control group membership. I use testing anxiety self-assessment, rather than state anxiety assessment, having proved to be the only relevant predictor of testing confidence (Model 21.2 see page 103). While state anxiety may also be associated with anxiety group membership, its impact on testing

confidence seemed negligible and would thus be an irrelevant confounding factor for the models above.

If testing anxiety is associated with anxiety group membership, randomisation 'failed' to create balanced groups, or more precisely, allowed that low anxiety individuals were sorted disproportionally into the anxiety group and, holding everything else constant, this group would appear more confident about political knowledge testing.

A Welch Two Sample t-test using the treated subsample of women suggest that this suspicion may be warranted ($t(98.15) = 1.98, p = 0.05$) where self-assessed anxiety is the scaled latent variable generated via an IRT model for use in Models 20.1 and 20.2 (Study 1 'Interim link 2'). Above-average anxiety women ($M = 0.27$ amounting to one-third of a standard deviation) were sorted into the control group while near-average or below ($M = -0.03$) into the anxiety group. There may be no comparable sorting effect for men ($t(104.19), 0.96, p = 0.34$) but the observed mean confidence in the anxiety group ($M = -0.14$)does appear lower than mean confidence in the control group ($M = -0.01$). Discarding the success of the emotion manipulation itself, this may be read as evidence that the anxiety group's higher-than-expected predicted testing confidence may be partially due to lower prior testing anxiety, at least for women.

Successful emotion manipulation is often checked with similar self-report of emotion following manipulation, however, due to limitations in survey length, the self-assessment questions were not re-asked after emotion manipulation. Instead, I am looking at the only available data directly relating to the less ambitious concept of 'compliance' with the writing task: *the written answers themselves.*

Dictionary-based methods are often used in quantitative text analysis to count the frequency of words and expressions relating to a concept defined as a list of relevant word stems or a particular combination of them. While the construction and basic validity of these dictionaries itself is an active area of research, especially the measurement of emotion in texts [141],[175], I am using here, rather heuristically, a large and relatively well established dictionary designed to measure psychological states in text, LIWC [140],[141]. LIWC defines 'affect' broadly through 915 word stems taken from 'common emotion rating scales such as the PANAS, Roget's Thesaurus, and standard English dictionaries' [176], from which 91 are meant to directly tap into anxiety (such as worr, fear, nervo etc.).

Using the frequency distribution of these word stems in the user-generated text response, each respondent gets a numerical proportion score that reflects the observed percentage of hits under the affect and anxiety categories. In practice, all respondents carried out two writing tasks, one when anxiety targets (or online shopping items) are listed and one when they selected one of these targets for further elaboration. Exploring these textual answers reveals that the list of anxiety targets, taken out of context, may themselves not convey anxiety:

"Money! My son living in London. My daughter staying out late"

Anonymous subject 1.

"Meeting new people. Ringing HMRC. Strange noise from car engine"
Anonymous subject 2.

but the follow-up text after each list may be more useful to pick up on affect and more specifically, anxiety:

"I worry about my son living and working in London due to the recent
terror attacks"
Anonymous subject 1.

"Meeting new people and having to talk to them. I can feel them assessing
me, and judging me, and although I know that I am every bit as good as they
are, I am almost guaranteed to say something stupid, or not know what they
are talking about. I feel they are looking down on me for the way I look,
dress and generally come across."
Anonymous subject 2.

Looking at women's textual answers, 'affective' word stems range between proportions of 0 and 16% in the control group, and 0 to 41.67 % in the anxiety group. There are no hits for the narrower category, 'anxiety' word stems, in the control group that wrote about items bought online. Hits under 'anxiety', however, range between 0 and 25% in the anxiety group. To illustrate these numbers using the two examples above: the first elaborate answer by Anonymous subject 1 is assessed to contain 18.75% of words with affective component and 12.50% falls specifically under anxiety, while only 3.03% affect in the second elaborate response by Anonymous subject 2. and 0% anxiety specific. This shows that the more explicit anxiety is on the level of diction, the better the software performance: Subject 2 has clearly complied with the anxiety task.

While admittedly rough measures, I used both affect and anxiety as word stem proportions as a proxy measure of affect and anxiety on the subject-level. Welch's Two Sample t-tests reveal differences in average affect between the anxiety group ($M = 7.16$) and the control group ($M = 1.37$, $t(146.6) = 9.18, p < 0.01$) as well as differences in average anxiety between the anxiety group ($M = 3.22$) and the control group ($M = 0.00, t(103) = 8.54, p < 0.01$) suggesting at least basic compliance with the emotion induction tasks, or more specifically, that the anxiety group used words and phrases explicitly relating to anxiety and a range of its synonyms.

An option would be to re-fit the previous models selecting only 'compliant' subjects using $> 0\%$ anxiety words. The second example above, however, demonstrates that compliant subjects managed to describe anxieties without using the expected vocabulary, thus exclusion may be unwarranted. And while I used emotion in text as a proxy of the in-

duced anxiety level, the two are not the same thing. Anxiety may be present in text but may not have enduring impact on state anxiety.

Further measures at this point are not readily available to test for the success of emotion manipulation. I will take up the suggestion about 'controlled behaviour' mentioned at the beginning of this section later in the conclusions.

5.2.4 Section Conclusions

My hypothesis was that political role models increase confidence in political knowledge via alleviating implicit anxieties stemming from negative stereotypes. In a 4 (ordinal picture treatment) by 2 (anxiety/control) experimental design, the expectation was that confidence would increase following the picture treatment only in the control groups, but flattens out in the anxiety groups. I introduced a few analytical steps to review the results and concluded that they do not follow this hypothesized mechanism, as the anxiety induction appeared to leave confidence unaffected and role model effects prevailed despite induced state anxiety.

Investigating this further, additional tests showed that the experimental anxiety group was composed of a greater number of low testing anxiety individuals, leaving the option open that state anxiety manipulation did not overwrite prior levels of testing anxiety. State anxiety itself may have been successfully induced: text analysis shows at least some basic level of its validity, participants did reflect on anxiety sufficiently. Based on my results, it seems possible that state anxiety does not relate to confidence in the way prescribed by a generalised theory of stereotype threat in politics, and more specific, situational anxiety matters, such as testing anxiety.

5.3 Study 3: Experimental affect and attentional bias

This study enables me to address some of the questions outside of the scope of the Studies 1 and 2. Drawing and expanding on stereotype threat theory, the studies above kept focused on anxiety, being central in previous literature. I did experiment with some degree of differentiation in anxiety type, to find no evidence of mediation using a state anxiety approach, but more evidence using a situation-specific, testing anxiety approach. This led me to further conclude that there is place for introducing more complexity in affect when examining responses to stereotype threat.

Another limitation was testing effects only on confidence about test answers as dependent variable. At the point of drawing conclusions in Chapter 4, I distinguished between (i) attention and election interest-related effects dominated by tokenism, in whereby a single (novel) woman's 'striking minority' presence may draw attention but not a greater presence of women; and (ii) confidence and self-efficacy related effects described by linear effects in which the greater women's numerical presence, the greater women's confidence and political efficacy. Studies 1 and 2 use the online experiment data, where I could only detect effects of the latter, linear type of role model effect on confidence and I was yet unable to test mediation related to tokenism.

In this study, I address these two points by going back to the lab experiment data where tokenism was first detected, and introduce a randomised mediation study on emotion drawing broadly on the framework of introduced in the theoretical framework, in which some emotions motivate approach and avoidance of threat, and introduced literature where extreme aversion or anger was a 'coping strategy' with stereotypes that did not result in a turn away from the source of stereotype threat but resulted in more engagement and confrontation. Based on this component of action, I hypothesised that anger may reverse political disengagement from male dominated politics, while the absence of this component: sadness and helplessness may result in disengagement (and perhaps more reliance on external 'help' such as role models).

5.3.1 Data & procedures

In the lab experiment, a small student sample of 28 women viewed pairs of pictures while wearing head-mounted eye tracker devices. 8 picture pairs, randomly presented 20 times in total, featured groups of politicians on the campaign trail, virtually identical except that women were cropped out in varying proportions, thus one of each pair featured a group with improved gender balance, the role model picture; and one with comparably worse gender balance, the contrast picture. The task was to identify a 'dot probe' replacing either type within each pair as quickly as possible with the aim of investigating attentional bias to role model pictures through performance on this task as well as through a number of measures generated by the eye tracker. More details on the experimental design are given in Chapter 3, and an extensive analysis of attentional bias to gender balance, broken down by specific numerical proportions, is presented Chapter 4 Section 4.1. Example screens featuring picture contrasts are shown in Figure 3.2 on page 3.2, the full pool of pictures is shown in Figure 3.3.

Randomisation inference is generated from the random shuffling of the 8 possible image pairs across trials and from the random presentation of the balance-improved picture on either the left or the right side of the computer screen. In this study, however, these manipulations are coupled with an experimental manipulation on emotion. Prior to picture presentation, participants were randomly assigned to reading-and-writing tasks connected to emotions experienced regarding everyday sexism. One group read and wrote about *anger*, another about *sadness and helplessness*, and a third, control group, read about product reviews and then described recently bought items online.

Chapter 2 Section 2.1.3 elaborates on the theoretical foundation behind the working hypotheses: that *cueing sadness* and a feeling of helplessness *motivates avoidance of the source of stereotype threat* because individual action to circumvent negative stereotypes seems infeasible. These participants may rather approach role models to avoid the source of stereotype resulting in attentional bias. On the other hand, *cueing anger may motivate approach of the source of stereotype threat* because individual action seems more likely, even unavoidable under extreme aversion, resulting in no attentional bias to role models. Thus emotion may help us understand more about heterogeneity in individual responses

to political stereotypes.

The result is a 2x3 design in which subjects' attention is measured on a (1) contrast image featuring unfavourable gender balance and on a (2) role model picture featuring favourable gender balance, broken down by the (a) anger, (b) sadness/helplessness, and (c) control (emotion irrelevant) groups.

The control group serves two purposes: a benchmark to evaluate emotion effects on those reported in Chapter 4, and more importantly, *similarities between either emotion groups and the control group may be suggestive of that emotion being relevant mediator of 'naturally observed' tokenism.* In other words, if cueing helplessness motivates approach of role models, then it is possible that effects on the 'helpless' may be key drivers of role model effects. Note, however, that a crucial comparison group is missing due to study limitations in a lab experiment: gender, as only women participants were recruited.

Table 5.10: Chapter 5 Study 3—Experimental design

| | | Picture pairs | |
		Contrast img. *'Men only'*	Role model img. *'Token'*
	Approach	Attention +	Attention +
Negative affect	Avoidance	Attention −	Attention +
	Control*	Attention −	Attention +

* This group has been analysed more thoroughly in Chapter 4 Section 4.1.

Hypothesized attentional bias across image pairs where Contrast image shows no women politicians and the Role model image has improved gender balance, in this study only Token presence (one woman or 20%) analysed.

Attention measures mostly follow those presented in Chapter 4. Dot probe task *response latency* variable shows, in milliseconds, the lag of probe detection, and has an empirical range of 57–997 ms. I am inferring an attentional bias to unfavourable or favourable gender balance pictures based on whichever received systematically quicker response times. Further attention measures are extracted from the eye tracker, single fixations—gaze kept on a single location—along with their duration in milliseconds are counted on each picture. Fixation durations range between 66.5 and 865.2 ms, and counts between 0 and 5. A greater *count* or *duration of* these *fixations* is read as evidence of attentional bias. Saccadic movement data, however, is discarded from this study for two reasons. Firstly, there was a strong similarity between duration and count measures across saccades and fixations, suggesting that these measures tap into similar attentional patterns strictly speaking in the context of my experimental design. Secondly, there were no clear hypotheses on saccades but my investigation in Chapter 4 was exploratory, looking at similar measures for cross-validation purposes rather than to learn about unique results. Thus to keep the study concise, I will not fit the emotion models on saccadic data as dependent measures.

As the part of the rationale behind this study was to investigate mediation of tokenism, explaining token effects on attention attention, I am discarding 4 picture pairs not featuring token presence of women but some other proportions. The remaining 4 picture pairs all contrast a token woman to men only politicians. I do, however, perform checks on the remaining groups and will report these as secondary findings.

After discounting trials that are not token versus men only pairings, incorrect submissions on the dot probe task, and missing data due to software error (poor tracking) or non-compliance with instructions to observe both pictures, I am working with the following sample sizes: $N = 261$ response latencies on the dot probe task nested within 27 subjects and $N = 368$ image-level observations across 184 trials (pairs presented, discounting practice trials) nested within 18 subjects in the eye-tracking data. A more detailed account on missing data management was introduced in Chapter 4 Section 4.1.

Similarly to Study 2, estimation is possible in a single step. There being only two picture treatment groups, and no three-way interaction to include participant gender, I am able to present the model directly and draw appropriate comparisons without the need to present intermediate models, as before. I continue using multilevel models fitted via MCMC to account for nesting within trials and participants on response latency and fixation duration dependent variables. Fixation count is modelled via multilevel Poisson regression fitted via MCMC.

5.3.2 Results

Table 5.11 presents group means descriptively. Control group means through all measures restate findings of attentional bias to tokenism reported extensively in Chapter 4: that pictures featuring women at this proportion receive quicker response times (and those response times have decreased variance), longer gaze duration, but only marginally greater number of fixations. Latter is consistent with studies documenting that in some cases, after attending to threat, disengagement proved difficult and attention remained fixed on targets [135].

The anger figures seem to support the hypothesis of approach: response times are down at levels comparable to control groups both for role model and contrast images, suggesting that the attentional bias disappears under anger. Further support comes from one of the eye tracker-generated measures as both pictures received increased gaze duration. There are no obvious effects on fixation counts but they do appear closer to each other than in the control group.

Sadness figures, finally, resemble those under control more than they do the anger predictions, confirming the hypothesis of avoidance, however, attentional bias seems much more modest, at least if measured through fixation duration. The gap appears sharp in probe response time, but contrast picture variance decreased compared with the control group.

The fitted models are Models 25, 26, and 27, for response latency, fixation duration, and fixation count, respectively. In these, 6 coefficients add up to the 6 group-level pre-

Table 5.11: Descriptive statistics: Respondent *N* & Attention measures.

Pairs feature 'Contrast' image with men & 'Role model' image with a token (=20–25%) woman

Affect	Picture	N	Mean	St.Dev.
Response latency probe detection (ms),				
total N = 261 test responses by 27 subjects.				
Control	Contrast	38	416.97	134.15
Control	Role model	29	378.21	89.15
Anger	Contrast	53	389.98	101.49
Anger	Role model	52	382.98	90.20
Sadness	Contrast	42	402.85	84.85
Sadness	Role model	47	380.57	87.37
Fixation duration (ms),				
total N = 368 trials by 18 subjects.				
Control	Contrast	55	197.84	151.31
Control	Role model	55	268.01	193.86
Anger	Contrast	80	226.69	231.44
Anger	Role model	80	226.27	215.67
Sadness	Contrast	49	248.55	190.80
Sadness	Role model	49	268.26	196.77
Fixation count				
total N = 368 trials by 18 subjects.				
Control	Contrast	55	1.24	0.92
Control	Role model	55	1.51	1.00
Anger	Contrast	80	1.20	1.06
Anger	Role model	80	1.23	1.04
Sadness	Contrast	49	1.31	1.04
Sadness	Role model	49	1.57	1.21

dictions as shown in the tables above (and an extra, seventh coefficient controls for probe location in gaze data to account for 'lagged' gaze from previous trials, as introduced in Chapter 4 in more detail, not used for making predictions).

Model-based predictions (Table 5.12) are largely confirmatory of these observations. The three figures below show the predicted response times, fixation durations, and fixation counts holding random effect variation constant as well as the (now weak or null) effect of probe location (an additional variable introduced in Chapter 4 to control for the possibility of 'lagged' gaze location due to probe presentation in the previous trial).

Looking at dot probe response latency (Model 25 and Figure 5.7), expressed in standard deviation terms, attentional bias to role model pictures is large in the control group (the distance between Intercept of .37 and 'Role mod pic.' Beta of -.42, over two thirds of a standard deviation) and translates into a predicted 40 milliseconds shorter response time on pictures featuring a token woman. By contrast, the coefficients associated with the impact of anger add up to an even attention, predicted at 379.86 and 384.69 milliseconds and visualised in the middle panel of Figure 5.7, to contrast and role model pictures, respectively. Compared across emotion groups, anger's effect may be a mixture of it decreasing response time to contrast pictures (-0.48), and increasing response times to role model pictures (+0.45) but these coefficients keep their negative or positive valence signs within wider credible intervals, 85% and 90%, respectively. There is a very small

probability of a sadness effect, the predicted Contrast response latency being decreased by about 20 milliseconds and the role model response latency virtually unaffected.
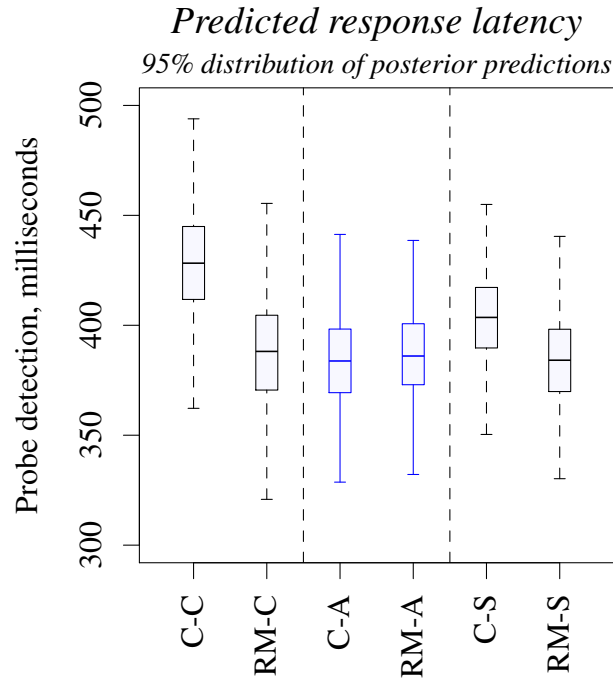
Note: **C-** and **RM** stand for **Contrast** and **role model** *(near 20% or token) pictures in picture pairs and* **-C**, **-A**, *and* **-S** *stand for* **Control** *(emotion irrelevant task),* **Anger** *('approach'), and* **Sadness/Helplessness** *('avoidance') groups. Faster probe detection may indicate attentional bias. For further details, see Table 5.12*

Figure 5.7: Attentional bias across picture contrasts, Model 25

The fixation duration model results (Model 26) are comparable to response latency results in that they are also expressed in the dependent variable's standard deviation terms (but inference changes as larger fixation duration, as opposed to shorter response time, is indicative of attentional bias). In the control group, the role model effect decreases somewhat to a third standard deviation, which is a weaker effect this time, but remaining positive within 90% credible intervals. The anger effect evened out attention where predictions are at 380 and 385 ms. Similarly to the dot-probe results, the posterior probability of the anger effect, increasing attention to Contrast pictures and decreasing attention to role models, decreases somewhat and remains around 85% also in the extended models. There are no sadness effects detected in these models either.

Finally, exploring Poission models on fixation count (Model 27), the model confirms that despite the very narrow range of the count variable, there may be a small marginal effect of the role model picture predicting, on average, a quarter more fixations but staying within non-zero positive 95% CIs. Under anger, the two picture types are predicted at 1.05 and 1.10 fixations where the probability of anger decreasing the count of fixations for role model pictures is 80%, while there is very little evidence that anger would increase fixations on the contrast images themselves. While these go parallel with the previous
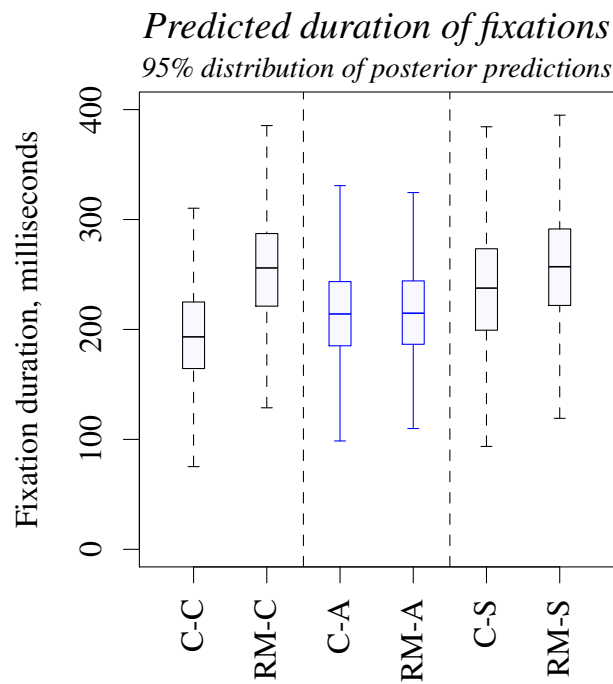
Predicted duration of fixations
95% distribution of posterior predictions

Note: **C-** *and* **RM** *stand for* **Contrast** *and* **role model** *(near 20% or token) pictures in picture pairs and* **-C**, **-A**, *and* **-S** *stand for* **Control** *(emotion irrelevant task),* **Anger** *('approach'), and* **Sadness/Helplessness** *('avoidance') groups. Longer fixations may imply attentional bias. For further details, see Table 5.12*

Figure 5.8: Attentional bias across picture contrasts, Model 26

models, there is a similar 80% probability of a sadness effect decreasing fixations on the role model picture suggesting that attention may be even under sadness as well.

5.3.3 Measurement validity

Similarly to Section 5.2.3 above, where I gave more details on the procedures, and another section on affect in text in Chapter 3 Section xx, here I report text analysis of written answers in the control, anger, and sadness groups using the LIWC dictionary.

The difference here, however, is that there is no meaningful inference due to low subject-level sample sizes. Through $N = 9$ Anger responses, affect averages at $M = 5.61$ per cent of total words, anger at $M = 2.82$ per cent, and no hits under sadness; $N = 7$ Sadness responses average $M = 5.36$ per cent of total words for affect, $M = 0.48$ per cent for anger, $M = 1.89$ per cent for sadness; and $N = 7$ Control responses (missing data is highest at Control, product review group perhaps not surprisingly) average at a surprisingly high $M = 4.93$ per cent of total words for affect, although $M = 4.28$ per cent of this is labelled 'positive emotion', no hits for anger, $M = 0.32$ per cent for sadness (due to one observation that used 3 per cent of total words sorted under sadness). Descriptively, this suggests at least compliance with the affect induction task.
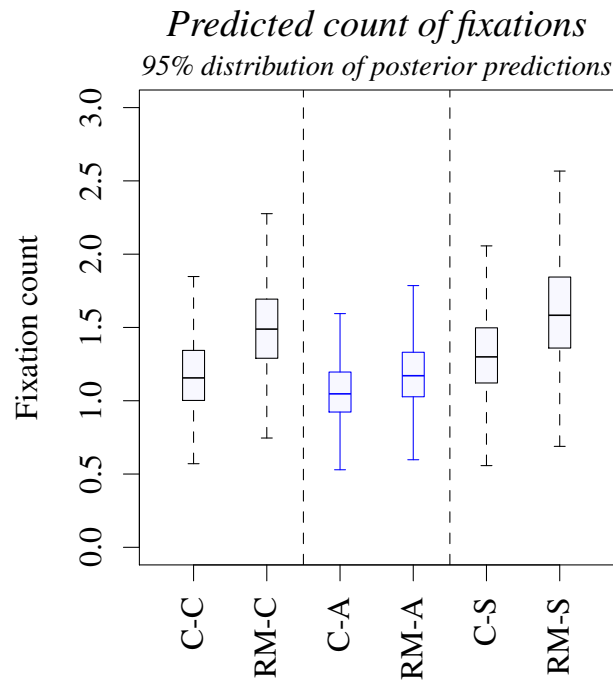
Predicted count of fixations

95% distribution of posterior predictions

*Note: **C-** and **RM** stand for **Contrast** and **role model** (near 20% or token) pictures in picture pairs and **-C**, **-A**, and **-S** stand for **Control** (emotion irrelevant task), **Anger** ('approach'), and **Sadness/Helplessness** ('avoidance') groups. Greater number of fixations may imply attentional bias. For further details, see Table 5.12*

Figure 5.9: Attentional bias across picture contrasts, Model 26

5.3.4 Section Conclusions

Affect and attentional bias

Through three measures of attention, in Study 3 I first restated that, without emotion manipulation, there is a natural tendency to direct attention to token presence of women as opposed to 'men only' political groups on a student sample of women, using manipulated campaign images from the 2015 General Elections.

I then proceeded to analyse this bias given experimental emotion: anger and sadness/helplessness, where the hypotheses were an approach of the source of stereotype threat in the anger group but avoidance in the sadness group. I thus evaluated the differences in predicted attention across contrast and role model pictures to infer on attentional bias, and within picture types I looked at emotion effects, to make appropriate comparisons with the control group.

An attention gap seems to disappear entirely under anger. This is often due to a mixture of increasing attention to Contrast, 'men only' image and decreasing attention to the role model image. Perhaps due to these simultaneous smaller movements, I did not have sufficient power in my tests to find conventional (i.e. 95%) certainty around these estimates but the probability of effects in the hypothesized, non-zero directions ranged between 80–85%. Given very similar movements across different measures of attention

Table 5.12: Models 25–27, Randomised affect study

	m25	m26	m27
DV	*Resp. latency, standardised (z)*	*Fixation duration, standardised (z)*	*Fixation count*
Type	Linear / MCMC	Linear / MCMC	Poisson / MCMC
Stacked *N*	261	368	368
Subject *N*	27	18	18
Trial *N*	20	20	20
DIC	684.52	989.11	984.65
	Beta [95% CI]	Beta [95% CI]	Beta [95% CI]
Intercept	0.37[-0.12;0.84]	-0.22[-0.70;0.29]	0.14[-0.29;0.56]
Anger	-0.47[-1.17;0.21]	0.11[-0.49;0.83]	-0.09[-0.66;0.46]
Sadness	-0.26[-0.89;0.44]	0.22[-0.54;0.94]	0.12[-0.59;0.69]
Role model pic.	-0.42[-0.88;0.02]	0.32[-0.01;0.72]	0.25[0.13;0.36]
Role m. x ang.	0.45[-0.07;1.04]	-0.32[-0.75;0.18]	-0.14[-0.30;-0.01]
Role m. x sad	0.22[-0.35;0.80]	-0.22[-0.70;0.28]	-0.05[-0.20;0.12]
Probe lag		0.09[-0.09;0.28]	-0.06[-0.12;-0.03]
	Gamma [95% CI]	Gamma [95% CI]	Gamma [95% CI]
Subject var	0.33[0.13;0.59]	0.27[0.07;0.51]	0.27[0.10;0.49]
Trial var	0.00[0.00;0.00]	0.00[0.00;0.00]	0.00[0.00;0.00]
Residual var	0.73[0.60;0.86]	0.82[0.70;0.94]	0.00[0.00;0.00]

(fixation duration, count, and performance on a dot-probe task), I do treat this as at least preliminary evidence that anger resulted in approach of contrast images.

Sadness shows very small movements compared to the Control group, with much small posterior probabilities, however, this also means that the token bias prevails. Posterior predictions are somewhat closer than in the Control group with their variances increased.

Finally, textual responses to the emotion manipulation task suggested compliance with the reading and writing task, although a small N of responses left no room for population-level inference. I found, however, that the dominating valence of emotion in the control group was *positive*, which may itself have implications for response and attention strategies.

Mediation of role model effects

As I outlined earlier, I can treat likeness of either emotion group's profile with the control group's profile as suggestive of mediation. Rather than treating it as a trade-off between anger and sadness, or approach and avoidance strategies, the purpose of this is to point at relevant emotions explaining tokenism, or perhaps excluding some.

In this framework, sadness emerges as another likely emotion describing tokenism, besides anxiety in the previous studies. It seems likely that, within a potentially rich range and mixture of emotions describing the experience of stereotype threat, sadness and the feeling of helplessness are emotions motivating avoidance of stereotyping political stimuli (or politics itself), and, at the same time, directing attention to role models, as external actors to resolve threat.

Anger, may not be to blame for withdrawal from politics, despite its relatively neg-

ative standing in political science literature. High aversion has been found in previous literature to describe how partisans feel about outgroup members, labelled 'expressive partisanship' [88]. Unlike anxiety, it has been found to decrease information search from a variety of sources [76],[177] but it was linked with mobilisation including protests and demonstrations [87],[178]. In my framework, anger motivates approach of threat, with the potential of more political engagement.

5.4 Chapter Conclusions

Exploiting three experimental designs, I found that a situation-specific approach to anxiety, namely 'testing anxiety', as well as other affect motivating withdrawal, namely sadness and helplessness, captured emotions under stereotype threat. Especially Study 2 excluded the possibility that a diffuse 'state anxiety', in which survey question nothing is 'at stake' explicitly but simply inquires emotion after treatment delivery, is a powerful mediator of role model effects. In my treatment, threat is relatively subtle, compared with the original stereotype threat paradigm where African American participants were explicitly told that their peers performed worse than White Americans. It was at the moment of testing—or its prospect, in the testing anxiety question—, where anxiety emerged, in turn explaining test answer confidence.

My studies further show the power of role models in two ways: firstly, in increasing proportions, women in politics alleviate situation-specific anxiety when it comes to assessing one's own skills (and, perhaps, efficacy in general); and secondly, my eye-tracking study on attention shows that at least a single standing, token woman may be sought out in contexts where one does not take action to confront stereotypes herself. By contrast, I found that in the presence of anger, women confronted stereotypes, and did not turn away from politics.

Notes

[22]Note that within the BES political efficacy questions, items about subjective assessments of political knowledge dominated as well.

[23]As explained in Chapter 3, participants randomly received either of the two questions, my intention being that one emotion is not projected over another.

[24]A third group carried out another task to reflect on women's efficacy in politics, analysed in Chapter 6, this group of people is excluded from analysis of this study.

[25]Attentional engagement with pictures of improved gender balance.

[26]Everyday sexism blog entries write up later.

[27]As well as to minimise the burden on test subjects, to keep the survey short and manageable within few minutes.

[28]As well as a policy 'Expectations lowered' group, participants not included here but in Chapter 6 Study 2 randomised mediation study, see Figure 3.4 on 40.

Chapter 6

Understanding the role model effect:
Expectations about policy competence and substantive representation

As much as I believe in the power of role models, [...] I suspect that the greatest contribution Obama will make to narrowing the achievement gap will be his policies, not his persona. [179]

There are, in the words of Theresa May, 'girl jobs and boy jobs'. She's a girl in a boy job. Make it count. [180]

This chapter investigates a causal mechanism in which women form more optimistic expectations about public policy with an increasing number of women representatives in government, a factor that motivates more political engagement instrumentally. Leveraging two experimental designs, I find that though such instrumental links exist, they are unlikely to mediate role model effects; if anything, those more optimistic about policy demonstrate lower self-efficacy.

Looking at literature on women's political representation, two linkages continue to dominate scholarly interest: whether 'descriptive' or numerical representation leads to better substantive representation in terms of public policy; and whether descriptive representation brings non-policy or 'symbolic' benefits to citizens such as a sense of empowerment, evidenced typically by greater political engagement. While the second linkage is of interest in my thesis, I outlined a causal mechanism that brings together both elements: *seeing women succeed* in office, in terms of delivering better or more relevant policy, citizens may become more optimistic or *confident* about their own voice or self-efficacy in the political domain. A similar mechanism has been proposed behind 'Black Empowerment' where Black voters see improved chances of realising self-interest through appointing Black executive leaders [22]. Correspondingly, Burns and colleagues hypothesise in-

strumental links between women in the electorate and women as political candidates as former expects latter to implement 'woman friendly' policies [17, p.352].

Empirical support for the full causal chain, down to the level of political engagement, is scant. The studies referenced above carry out no formal test of this mechanism. The first and only test of a relevant mechanism, a cross-sectional study on adolescent girls' anticipated political involvement measuring perceptions of government responsiveness, finds no evidence [19].

These theoretical and empirical approaches face a number of limitations. Firstly, existing theory is often restrictive in that it does not discuss motivations other than those of pure self-interest, or policy other than 'woman friendly'. Secondly, if taken seriously, the notion of a 'woman friendly' output implies that women prefer policies based on some degree of gender identification. At least in terms of voting behaviour, however, studies show that gender is rarely a relevant social identity driving choice [29],[181]. Finally, the attempt above to measure government responsiveness as a mediator may capture a holistic judgement of policy—but it may be more than that. Beliefs about government responsiveness, asked often as a single survey question, is likely to be less sensitive to expectations about individual candidates or the composition of legislatures but may influenced by a number of more stable individual attitudes, on the one hand, and citizens' 'baseline' psychological engagement with politics: the dependent variable itself. Thus even if detected, evidence of an effect on government responsiveness may expand the host of dependent variables, but may not reveal much about the underlying mechanisms.

In this chapter, I describe the results of two studies investigating how gender, and gender balance in political groups, relate to expectations about the collective competence of political groups in delivering good policy in general, and to a generalised notion of delivering women-friendly policies more specifically. In *Study 1*, survey respondents are invited to "rate how well groups of politicians would manage given issues". Group competence ratings gathered this way are then linked with the experimentally manipulated gender balance in picture stimuli on the one hand, and political engagement, on the other. In this classical, regression-based approach to mediation analysis, mirroring that of Chapter 5 Study 1 where the mediator variable is observed after treatment delivery, rather than itself experimentally manipulated, the hypothesis is that an improving gender balance leads to an increased political engagement because of improved group competency judgements.

Instead of defining women-friendly policies, I explore two key policies: healthcare and finance, understood generally and kept broad enough to accommodate self-interest or group-based interest, but specialist in the sense that stereotypes may prevail in how well suited politicians of either gender are in managing them [182][183]. I cite relevant literature about public views of competencies in Chapter 2 Section 2.2.1, showing that though men's and women's issues vary across contexts, depending, crucially, on the distribution of labour, public opinion in Western democracies tends to subscribe to issue stereotypes linking women politicians with 'compassion' issues such as healthcare, childcare, and

issues related with welfare provision. By contrast, many form associations automatically between men and economics, finances, military and defence. Furthermore, these areas are also meant to cue separate dimensions of policy making in terms of ideology, while finance may remind people of more conservative ideas of policy making in the context of a small state, healthcare may remind people of welfare provision in the context of the welfare state, argued to be more conductive to gender equality [184].

In *Study 2*, the crossover experimental approach, I seek confirmation that an increase in political engagement is in fact due to those improvements in expectations about policy, but this time specifically about women friendly policies, by introducing 'negative' manipulation on the mediator (see also Chapter 5 Study 2). By citing an accessibly written academic study casting doubt on whether women—by the sole virtue of being women—pursue different policy than men, my aim is to explicitly *remove or lower those expectations* hypothesized under role model effects. I thus obtain treated groups, viewing an increasing number of women in picture stimuli, who are hypothesized to be putting faith in female politicians enacting women friendly policies; as well as treated groups, also viewing an increasing number of women in pictures, but expectations explicitly removed in this regard. My study compares political engagement under both scenarios to make an improved causal inference on mediation. The next sections elaborate on these procedures and report findings.

It is important to note that, similarly to my mediation studies in Chapter 5, I am tied to study mediation of role model effect that were detected in the second round of survey experiments. Being unsuccessful in detecting effects on political interest, external efficacy, and trust, I am only able to test mediation on test answer confidence as a dependent variable. Looking at the core literature cited above, however, I can maintain that instrumentalism is relevant for confidence, if interpreted as part of self-efficacy in politics, in that holding optimistic views about the success of 'descriptive representatives', one might become more confident about her own skills and voice in the political domain. Though imperfect, test answer confidence is meant to be a proxy of political self-efficacy understood this way.

6.1 Study 1: Perceived group competence as mediator

6.1.1 Data & Procedures

Similarly to the first two studies on mediation looking at anxiety, see Chapter 5 Study 1 on page 94 and Chapter 5 Study 2 on page 94, I am now working with data collected through the second or 'replication' round of survey experiments, specifically designed to investigate the underlying mechanisms of role model effects. An $N = 667$ sample of men and women viewed and rated pictures of politicians depicted on the campaign trail, manipulated to feature either only men or 0%, or a token presence or near 20%, a critical mass presence or near 30%, or a parity presence of women or 50%. Each participant viewed a series of 5 different pictures that belonged to either of these gender split options,

that is, no one received picture stimuli where gender balance varied across these options.

While all studies thus far take advantage of randomised exposure to these images, this is the first that looks at how participants *rated* picture stimuli. Under each picture in a series of 5, participants were asked to submit a rating of 1 to 5 stars to indicate how well they thought the group would manage a given policy area. Two possible policy areas, finance and healthcare, were displayed randomly under the picture stimuli thus one participant could rate a particular picture on healthcare while another would rate the same picture on finance. The image pool is shown in Chapter 3, Figure 3.3 on page 39, and the question wording is shown in the Appendices (first in questionnaire).

Along with group-level sample sizes, Table 6.1 shows average ratings per treatment group and participants gender descriptively, averaged within the two policy areas. There are only a third as many participants assigned to 'men only' picture viewing and rating tasks than to the rest of the gender balance setups: the extra respondents in the treated groups other than 'men only' participated in additional randomised mediation studies reported in Chapter 5 Study 2, and later in Chapter 6 Study 2. When comparing average ratings across treatment groups, participants receiving those further manipulations can be included in the analysis as the randomised mediation tasks were carried out after picture rating. These participants will be, however, excluded when looking at measures taken *after* further manipulations, including confidence during the political knowledge test.

Table 6.1: Descriptive statistics: Sample sizes & Mean policy ratings.

Policy	Women pct	Gender	Resp. N	Stacked N	Mean rating (1–5)	St.Dev.
Finance	0	Men	37	86	3.03	1.11
	Token 20		92	252	3.01	1.11
	Cmass 30		100	257	3.11	1.02
	Parity 50		98	271	3.00	1.05
	0	Women	41	104	2.81	1.06
	Token 20		89	203	3.00	1.08
	Cmass 30		96	254	3.12	0.94
	Parity 50		95	245	3.14	1.00
Healthcare	0	Men	37	104	2.57	1.07
	Token 20		93	218	2.95	0.96
	Cmass 30		99	253	3.28	1.08
	Parity 50		98	243	2.90	1.09
	0	Women	38	101	2.61	1.09
	Token 20		89	252	3.12	0.99
	Cmass 30		94	236	3.17	1.02
	Parity 50		93	250	3.05	1.03

"*You'll be looking at five campaign pictures, all from the UK. Under the pictures, you will find an option to rate how well the politicians as a group would possibly manage given issues. The more stars you give the better job you think they would do. You are welcome to rely on your gut feelings or whatever impressions you may extract from the imagery*"

Looking at descriptive statistics, finance ratings submitted by men do not fluctuate across treatment groups, suggesting that gender composition was irrelevant for their perceptions of policy competence on the collective group level. This is not the case for women respondents, whose ratings of group competence gradually improve with the addition of more women. Gender balance seems to matter in perceptions about healthcare policy both for men and women, both rating group competence higher as gender balance

approaches 50% parity. Interestingly, baseline expectations are comparable in healthcare suggesting no persistent gender gaps in rating policy efficacy, whereas the gender gap is more prevalent if only men are seen to manage finance. A more thorough inspection of the distribution of ratings reveals that high scale points (four and five stars) and low scale points (one and two stars) are often equally distributed, suggesting that participants may not have distinguished them and I thus decided to recategorise my rating scales into simpler, low-medium-high ordinal scales.

The next sections follow an analytical path mirroring that in Chapter 5 Study 1, presented visually in 'Chapter 5 Study 1—Steps of mediation analysis' on page 97. The first section re-states the 'direct link' between treatment and effect, as demonstrated in Chapter 4 Section 4.2.2. (pooled samples through two data collection rounds) and Chapter 5 Section 5.1.2. (current, 'replication round' sample only), namely that gender balance in picture stimuli increased women's confidence in their answers on the political knowledge test, in both sections presented as part of political efficacy. Then I examine the first 'interim link' to mediation, that is, if men's and women's ratings of group-level competence in managing finance and healthcare policies follow gender balance seen in the picture stimuli. I then proceed to test the 'second interim link' to see if variations in participants' rating of group competence explain differences in test answer confidence. Finally, a fully mediated path regresses both treatment and mediator on testing confidence in which suggestive of mediation may be group competence ratings capturing treatment effects entirely, as found in Chapter 5 Study 1 with testing anxiety.

6.1.2 Results

Direct path: From treatment to effect

Chapter 4 Section 4.2.2. and Chapter 5 Section 5.1.2 report evidence behind a role model hypothesis on test answer confidence, namely that women gave consistently more confident answers on the 10-question political knowledge test as they viewed more women in the picture stimuli, replicated in two data collection rounds and different test types. For this study, the direct path model is identical to Model 18, reported in Chapter 5 Study 1, restated below.

Test answer confidence is measured after each answer on a 1-10 scale, and test questions in this round of data collection (used in further sections of this chapter) were about matching domestic and foreign leaders with their jobs with varying difficulty (the extremes being the leader of the Labour party, and the Prime Minister of Italy).

The data structure is hierarchical in that confidence ratings are nested within individuals as well as test questions. The fitted model is thus a multilevel MCMC regression with flat priors where the random effects terms are of test subjects (respondents) and test questions. The treatment effect is teased out via an interaction term of women's proportion in picture stimuli (0%, 20% token, 30% critical mass, 50% parity) and respondent gender. Results are plotted in ''Direct path' results, Model 18' on page 98 while the full model

Group competence in finance

95% distribution of posterior predictions
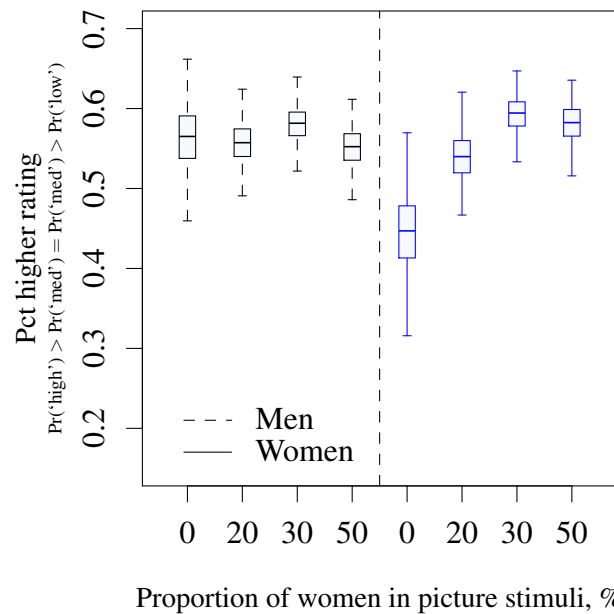
Proportion of women in picture stimuli, %

Group level predictions and 95% CIs at discrete proportions in picture stimuli: near 20% or token, near 30% or critical mass and 50% or parity.

Figure 6.1: 'Interim link 1' results, Model 28.1

results are in Table 6.3 on page 133.

Interim link 1: From treatment to mediator

This link explores the relationship between gender balance in picture stimuli, as independent variable and survey respondents' ratings of group competence in managing finance and healthcare policies, as dependent variables. Observed averages presented in Table 6.1 indicate that women, not men, rate group competence in finance higher as gender balance approaches parity, and that both genders will infer higher group competence in healthcare policy with more women on board.

The models, Model 28.1 (finance) and 28.2 (healthcare), are fitted on the ordinal, low–moderate–high group competence rating variables as multilevel ordered probit[29] regressions (via MCMC) where up to 5 picture ratings are nested within individuals (thus sample sizes vary according to how many pictures respondent rated, see group N's in Table 6.1), accounting for similarities in ratings given by the same individual. I use weakly informative 'Gelman priors' [185] recommended for better convergence of generalised linear models[30].

All model coefficients cumulatively designate a value on the z distribution, which can in turn easily be translated into a predicted probability that the dependent variable takes a 'higher' value on the rating scale, fixed across cutpoints meaning that it remains constant regardless whether looking at the probability of rating group competence 'moderate' rather than 'low', or 'high' rather than 'moderate'. A z score of 0 means 50-50% probabil-

Group competence in healthcare

95% distribution of posterior predictions
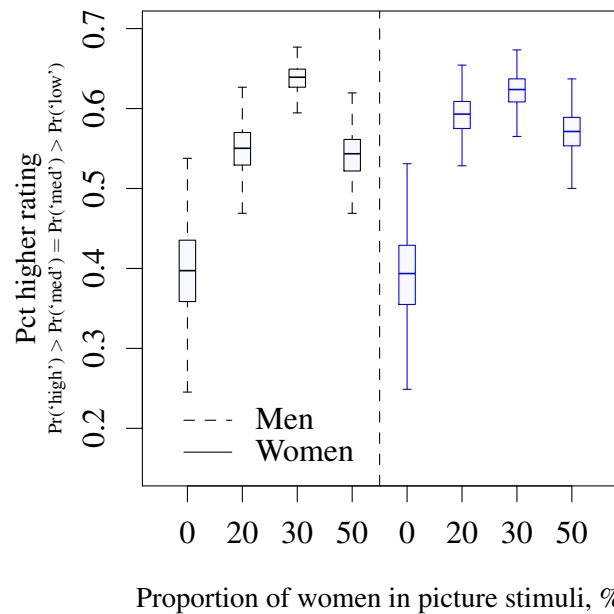
Group level predictions and 95% CIs at discrete proportions in picture stimuli: near 20% or token, near 30% or critical mass and 50% or parity.

Figure 6.2: 'Interim link 1' results, Model 28.2

ity of either low or high ratings. Positive coefficients will predict increasing probability of higher ratings across any cutpoint, and negative coefficients predict increasing probability of lower ratings across any cutpoint.

Looking at the finance model results, the intercept of $\beta = .84$ predicts that relatively high finance ratings are likely in the baseline category, namely those submitted by men looking at only male politicians (predicted probability shown in Figure 6.1). Treatment coefficients are suggestive that men's ratings of group competence in finance do not change as a function of gender balance in picture stimuli. The single standing coefficient associated with ratings submitted by women is negative within 95% credible intervals, lending high certainty to the hypothesis that women rate group competence in finance lower if group members are all men, with a marginal effect corresponding to -12%. The interaction effects show that those ratings will, however, increase by parity. The predicted percentages of women submitting a higher group competence rating are as follows: 45% in the men only condition, 54% in 20% token, 59% in 30% critical mass, and 58% in parity, thus average treatment effects go up to +13%, about evening out the initial gender gap.

Treatment effects look different in the healthcare model. Both men and women get *z* scores nearer the 50-50 cutpoint (in other words, the intercept is closer to 0) when looking at picture stimuli with only male politicians suggesting that ratings are not consistently high. Treatment effects are all positive: with the addition of a token woman in picture stimuli, group competence in healthcare policy is rated better. The wide credible intervals

Confidence about test answers

95% distribution of posterior predictions
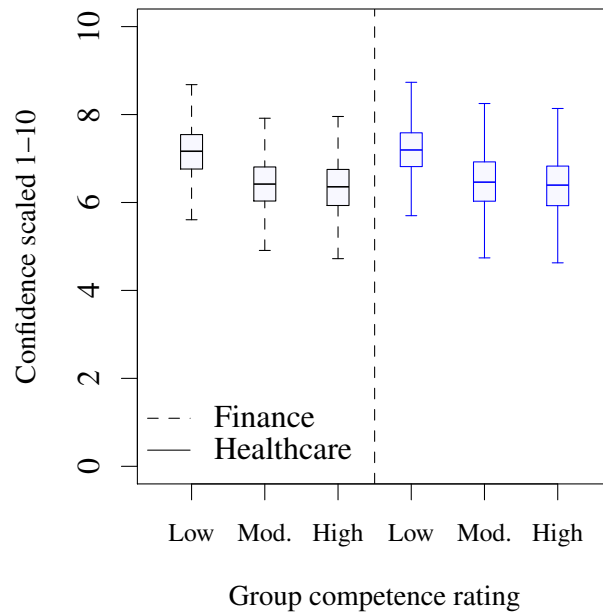
Predictions and 95% CIs at 'high', 'moderate', and 'low' group competence ratings across two policy areas, measured on five pictures featuring groups of politicians on the campaign trail.

Figure 6.3: 'Interim link 2' results, Models 29.1 (left) & 29.2 (right panel)

around the interaction effects suggest that this pattern may not get reversed for women respondents either. Figure 6.2 visualises the predicted probability of a higher rating, for both gender, starting at approximately 40% in the men only condition, peak above 60% by critical mass, and decrease only a few percentage points by parity.

Both models were unable to calculate a residual variance component, but they detected larger subject-level variation in the healthcare model than in the finance model. Full results are presented in the first two columns of Table 6.2.

Interim link 2: From mediator to effect

The second interim link tests directly for effects of group competence ratings on testing confidence, discussed within the group of indicators relating to self-efficacy.

Descriptively, confidence levels averaged within low, moderate, and high group competence ratings, suggest that *higher* perceptions of group competence predict, if anything, *lower* levels of confidence, $M_{low} = 7.14$, $M_{mod} = 6.44$, $M_{high} = 6.36$; contrary to the hypothesis of mediation predicting high self-efficacy once politics itself seems more efficacious with better instrumental reasons to engage with.

Models 29.1 (finance) and 29.2 (healthcare) provide confirmation of this inverse mechanism. Both models use test answer confidence as a dependent variable, nested within respondents and questions. The intercept may be interpreted as the expected value of test answer confidence if group competence is rated low, in both models near 7 on a 10-point scale. Two dichotomous variables identify the effect of 'moderate' and 'high' ratings

against this baseline. In both the finance and the healthcare models, these higher ratings translate into drops in confidence, $\beta = -0.73$ and $\beta = -0.80$ on the 10 point scale, small effects but comfortably within 95% credible intervals. Figure 6.3 shows the posterior distribution of these predictions. Table 6.2 shows the complete model results.

Table 6.2: Models 28.1–28.2 and 29.1–29.2, 'Interim links'

	m28.1	m28.2	m29.1	m29.2
DV	*Competence in finance* Pr(Y) = Higher	*Competence in healthc.* Pr(Y) = Higher	*Confidence in test answer, scaled 1–10*	
Type	Ordered probit / MCMC	Ordered probit / MCMC	Linear / MCMC	Linear / MCMC
Stacked *N*	1673	1657	2700	2650
Subject *N*	648	641	270	265
Item *N*	3	3	10	10
DIC	3368.19	3205.00	12.691	
	Beta [95% CI]	Beta [95% CI]	Beta [95% CI]	Beta [95% CI]
Intercept	0.84[0.48;1.25]	0.15[-0.24;0.53]	7.16[5.76;8.38]	7.19[5.97;8.47]
Gender: Woman	-0.51[-1.00;0.02]	-0.01[-0.53;0.57]		
Treatment: 20%	-0.04[-0.51;0.39]	0.62[0.15;1.12]		
Treatment: 30%	0.08[-0.36;0.51]	1.15[0.66;1.61]		
Treatment: 50%	-0.07[-0.50;0.39]	0.58[0.14;1.09]		
Woman x 20%	0.43[-0.17;1.02]	0.23[-0.43;0.87]		
Woman x 30%	0.58[0.00;1.16]	-0.10[-0.73;0.55]		
Woman x 50%	0.66[0.09;1.21]	0.16[-0.43;0.85]		
Rating: moderate		-0.73[-1.33;-0.10]	-0.75[-1.44;-0.15]	
Rating: high				-0.80[-1.57;0.00]
	Gamma [95% CI]	Gamma [95% CI]	Gamma [95% CI]	Gamma [95% CI]
Subject var	0.38[0.19;0.58]	0.71[0.46;1.03]	4.11[3.34;4.97]	4.12[3.22;4.96]
Item var			3.44[0.92;7.66]	3.34[0.81;7.28]
Residual var	1.00[1.00;1.00]	1.00[1.00;1.00]	5.88[5.57;6.21]	5.87[5.52;6.18]
	Theta [95% CI]	Theta [95% CI]		
Cutpoint	1.45[1.34;1.56]	1.51[1.40;1.63]		

Two follow up analyses confirm that firstly, this mechanism stands after accounting for respondent gender, with the magnitude of the fixed effect predictors somewhat increased ($\beta = -0.78$ of moderate ratings, and $\beta = -0.91$ of high ratings), and secondly, this mechanism does not differ across the two genders. Regardless of policy area, the interaction effect of moderate ratings with gender is estimated to be zero. There may be an interaction effect of high ratings with gender, suggesting that unlike men's, women's confidence did not decrease while giving high competence ratings, but these coefficients have very high posterior variance thus unlikely to provide a convincing model of test answer confidence.

Fully mediated path

The fully mediated models fit treatment and mediator (Model 30.1 finance, and Model 30.2, healthcare) effects simultaneously on testing confidence, taking the same functional form as the direct path but extended with the group competence rating variables, two dichotomous variables showing the effect of 'moderate' and the effect of 'high' ratings against 'low' rating.

In these extended models, negative effects of perceived group competence on testing confidence gain magnitude and both moderate and high ratings now correspond to one point (a third standard deviation) decrease on the confidence scale, regardless of whether the mediator variable measures finance or healthcare policy ratings.

Treatment effects prevail and the group estimates gain precision in terms of posterior variance, instead of disappearing altogether if mediation was detected. Identical to the direct path model, the key finding remains that men's testing confidence decreased, and women's confidence increased when viewing 50% parity images, as opposed to their baseline confidence when viewing men only groups. The predictions further suggest that men's confidence benefited from a slight boost, and women's confidence decreased slightly under token presence of women politicians suggesting a sense of curvilinearity in role model effects.

Jointly, these two movements do not support the hypothesis of mediation. Demonstrated by the results of Study 1 in Chapter 5, the regression-based approach would treat vanishing treatment effects with enduring mediator effects as evidence that expectations about group level competence in at least two domains would mediate role model effects on confidence. By contrast, I obtained some *parallel* effect of treatment and mediator in which the two, improving gender balance in politics and perceptions of group competence, have opposing effects on confidence and perhaps on self-efficacy in general.

The magnitudes of these effects are comparable. The role model effect in parity appears to be only somewhat larger than that of group competency ratings by about a quarter point: while women's average predicted confidence increases from 6.32 to 7.55, the average predicted decrease in confidence as ratings get higher amounts to about 1 point.

Table 6.3: Models 18 & 30.1–30.2, Direct and mediated ('full') paths

	m18	m30.1	m30.2
DV	*Confidence in test answer, scaled 1–10*		
Type	Linear / MCMC	Linear / MCMC	Linear / MCMC
Stacked N	2760	2700	2650
Subject N	276	270	265
Item N	10	10	10
DIC	12958.73	3368.19	3205.00
	Beta [95% CI]	Beta [95% CI]	Beta [95% CI]
Intercept	7.17[5.93;8.68]	7.88[6.64;9.31]	7.92[6.49;9.25]
Gender: Woman	-1.42[-2.27;-0.48]	-1.56[-2.35;-0.61]	-1.60[-2.58;-0.67]
Mediator: 'moderate' rating		-1.02[-1.61;-0.32]	-0.96[-1.51;-0.36]
Mediator: 'high' rating		-1.13[-1.86;-0.44]	-1.08[-1.76;-0.38]
Treatment: 20%	0.68[-0.21;1.61]	0.88[0.08;1.86]	0.79[-0.19;1.71]
Treatment: 30%	0.18[-0.73;1.10]	0.34[-0.60;1.27]	0.22[-0.72;1.21]
Treatment: 50%	-1.13[-2.08;-0.24]	-1.26[-2.24;-0.22]	-1.12[-2.06;-0.19]
Woman x 20%	-0.74[-2.04;0.60]	-0.60[-1.93;0.61]	-0.54[-1.91;0.82]
Woman x 30%	0.17[-1.16;1.57]	0.26[-1.14;1.58]	0.34[1.05;1.68]
Woman x 50%	2.17[0.88;3.40]	2.49[1.26;3.88]	2.36[1.09;3.68]
	Gamma [95% CI]	Gamma [95% CI]	Gamma [95% CI]
Subject var	3.68[2.97;4.45]	3.57[2.91;4.32]	3.55[2.82;4.24]
Item var	3.47[0.99;7.17]	3.50[1.05;7.59]	3.34[0.77;6.99]
Residual var	5.86[5.53;6.20]	5.88[5.57;6.21]	5.87[5.53;6.21]

6.1.3 Section Conclusions

This study investigated mediation of role model effects on women's psychological engagement with politics, more specifically confidence in political knowledge, introduced in Chapter 4 within a host of measures tapping into one's self-efficacy in politics. The working hypothesis was that women in visible positions may be perceived to implement better policies, contributing via a more diverse set of competencies, the success of role models positively impacting one's self-efficacy.

In a survey experiment, the three key concepts were operationalised as follows: treatment was experimentally manipulated gender balance in picture stimuli featuring political campaign groups; the mediator was perceived group competence, provided via ratings under each picture, in implementing finance or healthcare policies; and the effect was test answer confidence, measured through a 10-question political knowledge quiz, each question followed by a rating of confidence about the answer.

Mediation

Overall, my results do not support the hypothesis of mediation as laid out in Chapter 2 Section 2.2, 'Policy competence'. When respondents judged groups to do better on either finance or healthcare, their own confidence in political knowledge decreased. This suggests there may be a *trade-off between group-efficacy and self-efficacy* in politics, in which skills or knowledge about governance attributed to (female) politicians as external actors decreases one's self-evaluation on those skills and knowledge.

Further tests confirmed that this mechanism may not depend on gender, suggestive of a psychological mechanism not related to (perceptions about) descriptive representation. There is weak evidence that women's confidence is not implicated when they give higher group-level competence ratings as much as men's confidence, however, there is no evidence that their confidence would increase. Furthermore, averaged across the two policy domains, the effect appears 'linear' meaning that in the higher the group competence rating, the lower the confidence rating. This is thus a potential area of future research in political psychology, with or without an emphasis on gender.

Expectations about substantive representation

Perceptions about group-level competence in executing policy does, however, relate to gender balance, and only partially follows issue stereotypes. My models find evidence that women judge group competence higher when gender balance improves, regardless of policy area. A token presence was enough to improve competence judgements significantly, but further, smaller improvements happened beyond token, peaking at critical mass proportion. Note that women's policy competence judgement in finance goes against gender–issue stereotypes, confirming a working hypothesis laid out in Chapter 2 stating that, with more female politicians on board, women become more optimistic about policy outcomes in areas where issue stereotypes would traditionally disadvantage (and exclude) them.

Men mirrored women's judgements on healthcare, but they judged group level competence in finance with no regard to gender, rating it high already when men were present. Thus men seem to take cues from politician's gender to judge policy, following issue stereotypes: rating a policy area associated with women's competence better if female politicians seem to contribute, but any proportion of men's presence was enough to convince them of good finance policy.

My results are suggestive, if not explicit, of an instrumental link between women in the electorate and female politicians. These group competence ratings are consistent with the view that women expect better policies from women, whether that's based on a shared notion of a 'woman-friendly' policy or entirely up to individual taste.

Furthermore, this could possibly link back to psychological engagement with politics through an indicator other than self-efficacy in politics. Positive feedback to policy may indicate greater satisfaction with governance, which has been linked causally to trust with democracy [186]. These options were, however, not ready for exploration as the current round of data collection was unsuccessful in detecting effects on trust, See Chapter 4 Section 4.2.4.

6.2 Study 2: Expectations about women-friendly policy

6.2.1 Data & Procedures

In this study, I am able to use a larger sample of survey respondents who participated in the randomised mediation study. An $N = 471$ sample of men and women with near-equal gender split received picture treatment identical to those reported above: respondents viewed (and rated) pictures of political campaign groups where gender balance was manipulated to feature only men, or a token/20%, critical mass/30%, or parity/50% presence of women politicians. To simplify, latter three groups are referred to as the 'treated' sample of respondents.

Before taking dependent measures of political engagement, treated participants were reshuffled in another round of randomisation, relating to the mediator variable. According to the hypothesis of mediation through better policy expectations, women who viewed female politicians are more optimistic about the course of policy and thus show increased political self-efficacy. The current study *removes* these expectations for a randomly assigned half of the treated sample by citing an accessibly written study that argues for no difference between men and women politicians in terms of policy output:

> *Please take a moment to think about the following two findings taken from a recent expert study about women in politics in the UK.*
>
> The notion of women's feminizing of politics is controversial [...] implying that as a group, they inherently share a set of policy preferences or ideology. Women are not a monolithic entity with a collective set of interests

and beliefs. One could cite many examples of female political figures who represent the antithesis of what many regard as "women's interests" (e.g., Margaret Thatcher). Representation [...] is not a guarantee that women representatives will act for women. [...] In finance debates, findings show that the most significant [factors in] debate participation are relevant committee membership, Conservative party [membership], and seniority, holding other explanatory variables at constant. Sex was not significant.

This mirrors negative manipulation in Chapter 5 Study 2 where anxiety is re-induced by participants having to first think about things making them anxious, with the important difference that there was no text to read about anxiety[31]. Here the text is meant to help respondents think about the policy output of women and men, the topic being less easily accessible than personal anxieties.

The hypothesis of mediation in this case is tested similarly to that in Chapter 5 Study 2: improved policy expectations may be responsible for women's increased testing confidence if it remains low also when those expectations are experimentally lowered. Thus women's testing confidence is expected to remain flat even if they viewed women in the picture stimuli. There are no explicit hypotheses relating to men, whose confidence has previously been shown to decrease when pictures featured more women.

To finish, participants carried out a writing task similar to the one in the anxiety study. Here they had to briefly reflect on what the main argument was in the excerpt. Participants in the control group, as well as all participants that viewed pictures with only male politicians in them, had to think about and reflect on grocery items recently bought.

Once they completed these tasks, respondents proceeded to answer the political knowledge test familiar from the studies above, pairing political office with UK and international politicians. The dependent variable of this study is self-assessed confidence on the political knowledge test, rated after each of the 10 test questions.

Table 6.4 below shows pooled, average confidence by treatment and mediation groups. Confidence in the control groups writing about groceries instead of women in politics was analysed already in Section 4.2.2 and restated in Sections 5.1.2. and 6.1.2. above.

The descriptive findings do not support the mechanism regarding mediation: women's expectation-lowered averages do not remain flat, on 'men only' levels but increase along with control group averages, suggesting that role model effects prevailed. In all cases, expectation-lowered averages are above control group averages. Interestingly, men's expectation-lowered averages do remain flat, comparable to their 'men only' exposure levels. This suggests that perhaps men's decrease in testing confidence is due to expectations about women pursuing different policy, once in office.

The models below tease out these group averages accounting for the hierarchical structure of the data, in multilevel regressions fitted via MCMC, where 4,710 confidence ratings are nested within 471 individuals and 10 political knowledge test questions of varying difficulty. To make sure appropriate comparisons are not missed, I am presenting

Table 6.4: Descriptive statistics: Sample sizes & Mean confidence ratings.

Women pct	Mediator	Gender	Resp. N	Stacked N	Mean conf.(1–10)	St.Dev.
0	Control	Men	38	380	7.13	3.47
Token 20	Control		36	360	7.83	3.06
20	*'Expectation lowered'*		24	240	7.41	3.17
Cmass 30	Control		35	350	7.33	3.22
30	*'Expectation lowered'*		34	340	7.52	3.11
Parity 50	Control		31	310	6.04	3.64
50	*'Expectation lowered'*		39	390	7.30	3.30
0	Control	Women	41	410	5.72	3.65
Token 20	Control		31	310	5.68	3.60
20	*'Expectation lowered'*		29	290	6.13	3.63
Cmass 30	Control		28	280	6.08	3.58
30	*'Expectation lowered'*		36	360	6.60	3.42
Parity 50	Control		36	360	6.76	3.28
50	*'Expectation lowered'*		30	300	6.92	3.38

these results step by step: the first two models estimate mediator effects on testing confidence using the treated subsample only, with and without an interaction with respondent gender; then I derive treatment x mediator group averages for men and women, separately.

6.2.2 Results

The first model regresses the mediator on testing confidence of all *treated* respondents that is, men and women viewing pictures with at least a 20% or token presence of women in them. The key parameter estimate in Model 31, presented in Table 6.5, is of 'Mediator: Expect.lwrd' finding no evidence for the hypothesis that removing the expectation about 'women acting for women' decreases performance on the political knowledge test (or self-efficacy in politics in general). The coefficient itself is positive, consistent with descriptive findings—and with Study 1 findings, see Chapter Conclusions—but a mere quarter of confidence point with large posterior variance.

The second model, Model 32 refines this analysis by conditioning mediator effects on gender, failing to detect further evidence. Comparing the single standing and the interacted mediator effects (interpreted as added differential if respondents are women) shows no or negligible differences in mediator effects across the two genders. There are no notable differences between variance components either except somewhat smaller individual-level variation, likely due to baseline gender differences in testing confidence, absent from the previous model.

Models 33.1 and 33.2 are identical in specification, aiming to tease out mediator effects *jointly with specific treatment effects* (i.e. gender balance in pictures), compared to those participants who viewed men only picture stimuli. A problem introduced in Chapter 5 Study 2 is that a model specifying an interaction with treatment and mediator would be 'rank deficient' meaning not all combinations of treatment and mediator categories are populated—those receiving men only picture stimuli were always in the control group, see also Chapter 3 Section 3.1.2. Thus groups of treatment and mediator combinations are created by hand, resulting in 7 groups:

1. men only & control,

Confidence about test answers

95% distribution of posterior predictions

Proportion of women in picture stimuli
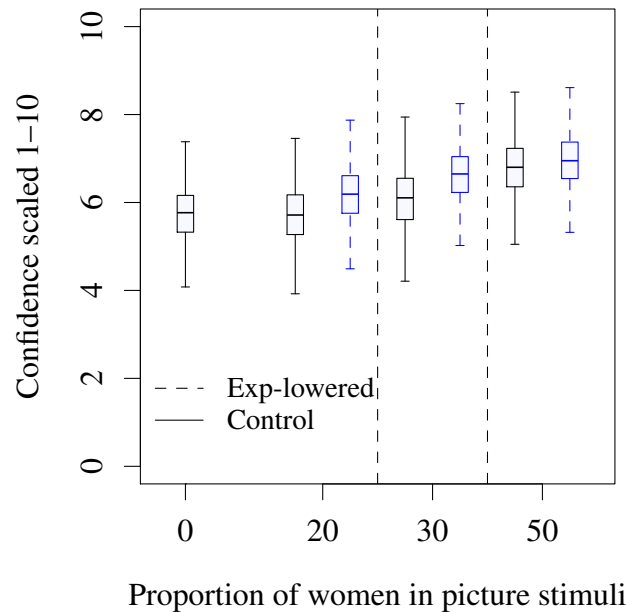
Women subsample only. Group-level predictions and 95% CIs at discrete proportions in picture stimuli: near 20% or token, near 30% or critical mass and 50% or parity.

Figure 6.4: Randomised mediation study, Model 33.1

2. token or 20% & control,

3. token or 20% & exp. lowered,

4. critical mass or 30% & control,

5. critical mass or 30% & exp. lowered,

6. parity or 50% & control,

7. parity or 50% & exp. lowered.

Keeping it as a categorical predictor, each coefficient compares the effect of group membership being (2) to (7) to the baseline, (1) men only control group, the estimate of latter being the Intercept. Model 33.1 uses the subsample of women, and Model 33.2 uses the subsample of men.

Looking at women, evidence consistent with the mediation hypothesis would be if any of treatment + 'exp.lwrd' coefficients were near zero (meaning no deviation from the baseline 0% men only exposure), paired with positive treatment + 'ctrl' coefficients, which would be the role model effect on subjects whose woman-friendly policy expectations were not removed. This is not the case. Parity effects prevail whether or not the mediator is introduced, the estimate in the mediation group being somewhat higher than in the control group ($\beta = 1.21$ and $\beta = 1.05$, respectively). These results are plotted in Figure 6.4 in terms of predicted probabilities.

Looking at men, whose average parity effect was a 1 point decrease in confidence, the negative manipulation on the mediator did have an effect, namely it *reversed* the parity effect: average confidence being not any longer different from that under men only picture stimuli. Complete model results are in Table 6.5).

Table 6.5: Models 31–32 and 33.1–33.2, Randomised mediation study

beta	m31	m32	m33.1	m33.2
subsample	Treated women	Treated women	All women	All men
DV		*Confidence in test answer, scaled 1–10*		
Type	Linear / MCMC	Linear / MCMC	Linear / MCMC	Linear / MCMC
Stacked *N*	3920	3890	2310	2370
Subject *N*	392	389	231	237
Item *N*	10	10	10	10
DIC	18249.10	18118.69	11002.18	10828.38
	Beta [95% CI]	Beta [95% CI]	Beta [95% CI]	Beta [95% CI]
Intercept	6.70[5.68;7.85]	7.09[5.99;8.33]	5.74[4.336.92]	7.13[5.70;8.47]
Mediator: 'Exp.lwrd'	0.28[-0.08;0.69]	0.28[-0.26;0.94]		
Gender: Woman		-0.89[-1.50;-0.30]		
'Exp.lwrd' x Woman		-0.07[-0.77;0.84]		
Group 20 & Ctrl.			-0.02[-0.93;1.02]	0.70[-0.07;1.66]
Group 20 & 'Exp.lwrd'			0.44[-0.59;1.42]	0.27[-0.71;1.21]
Group 30 & Ctrl.			0.35[-0.73;1.37]	0.23[-0.65;1.13]
Group 30 & 'Exp.lwrd'			0.90[-0.01;1.81]	0.39[-0.50;1.32]
Group 50 & Ctrl.			1.05[0.09;1.98]	-1.07[-1.98;-0.19]
Group 50 & 'Exp.lwrd'			1.21[0.26;2.16]	0.17[-0.65;0.97]
	Gamma [95% CI]	Gamma [95% CI]	Gamma [95% CI]	Gamma [95% CI]
Subject var	4.03[3.45;4.71]	3.76[3.17;4.38]	4.00[3.18;4.87]	3.27[2.58;3.95]
Item var	3.01[0.84;6.63]	2.96[0.92;6.08]	3.13[0.82;6.60]	3.49[1.14;7.40]
Residual var	5.62[5.34;5.88]	5.64[5.37;5.94]	6.25[5.87;6.64]	5.16[4.84;5.46]

6.2.3 Section Conclusions

This study exploited another crossover experimental approach to investigate mediation of role model effects on women's confidence about their political knowledge, discussed earlier as part of self-efficacy in politics, testing the hypothesis that female politicians are perceived to implement 'woman-friendly' policies.

Treatment was randomisation of gender balance in picture stimuli featuring groups of politicians on the campaign trail. The dependent variable, confidence in test answers, was measured on a 10-question political knowledge quiz after each test question. According to the hypothesis, women were expected to form more optimistic expectations about the course of policy, thus the *negative manipulation* on the mediator concerned the removal of these expectations. Participants read about, and reflected on theoretical arguments outlining why women would *not* act on behalf of other women, once in office. In the control group, they wrote about grocery shopping.

My models tested, by making a number of comparisons, if testing confidence decreased in this expectation-lowered group compared to the control groups. I found no evidence of an overall difference between expectation-lowered and control group confidence,

averaged across gender balance scenarios. Small differences pointed to the direction opposite to the hypothesis. Looking at group averages broken down by gender balance in picture stimuli, however, I found that although role model effects still prevail for women, the mediator manipulation did impact men's confidence. Though not hypothesized explicitly, this suggests that men lose confidence (and perhaps political self-efficacy) only if women in office are expected to act for fellow women in the electorate. If that default expectation is removed, men remain confident (and perhaps efficacious).

This approach to mediation, evidently, faces a number of limitations. While the component of randomisation improves inference, it is difficult to manipulate a hypothesized psychological mechanism in a strictly controlled way. In this vein, the first limitation is that *a check on compliance is difficult to carry out*. While the textual responses are available to analyse in terms of grammatical structure, even sentiment and elements of cognition via dictionary-based methods, such dictionaries evaluating if respondents have adopted a particular view about the subject (i.e. women carrying out policies to help other women) would be more difficult to build. We cannot be sure that participants reading and reflecting on the text have lower expectations regarding feminist policy than those in the control group.

A second limitation is that behaviour measured after manipulation on the mediator may not be directly due to changes in expectations about policy, but something else, such as cueing the relevance of gender in politics, or maybe the fact that these participants were reminded to think about policy output (regardless if by men or women) in the first place while the control group was left to follow implicit evaluations.

Finally, this study also faces limitations given the specificity of the dependent variable, testing confidence: it is possible that effects would be detectable on other indicators of political engagement with a more explicit focus on policy expectations, such as satisfaction with governance or external efficacy. In this round of data collection, however, I was unable to detect role model effects on these variables, see Chapter 4 Section 2.

6.3 Chapter Conclusions

Chapter 6 investigated mediation of role model effects through a more instrumental mechanism, namely that women's political engagement is driven by expectations about the efficacy of politics itself, improving when more women hold visible political office. My studies looked at this idea first through broad policy areas, of finance and healthcare, looking for holistic judgements about group competence in groups of politicians as a function of their gender balance. Aiming to provide more detail on this broad mechanism, a second study looked at specifically whether women form expectations about female politicians enacting 'woman friendly' policy.

Both studies are based on a survey experiment where participants viewed and rated a series of pictures featuring 2015 campaign groups of unknown politicians with gender balance manipulated to feature only men, a token or 20% presence, a critical mass or 30% presence, or a parity or 50% presence of women. Study 1 analysed the rating of these

groups on 'how well [they] would manage finance/healthcare policy', and linked this to the gender balance in picture stimuli on one hand, and performance on the 10-question political knowledge test on the other. The key dependent variable was participants self-assessed confidence after each test answer given on a 10-point scale, thus I analysed an $N = 2,700$ confidence ratings nested within 270 respondents with near equal gender split, and within 10 test questions.

The second study introduced randomisation on the mediator itself, and motivated a randomly selected sample of people to consider that women may *not* act for other women as policy makers. Using the same dependent variable, here I analyse an $N = 4,680$ confidence ratings, accounting for the same nesting as a function of gender balance in picture treatment and across the mediator (where expectation is lowered) and control groups. In this setup, women viewing more women in picture stimuli are hypothesized to be more politically 'efficacious' because of specific expectations from women in politics *if the expectation-lowered subsample give less confident answers*, despite viewing more women in picture stimuli.

Based on these studies, I draw a number of conclusions about mediation and connected mechanisms as summarised below.

Mediation through instrumental links

While theoretical work has used instrumental considerations as a point of departure linking descriptive representation and political engagement [22] including psychological engagement with politics [30], there is as yet no empirical evidence of instrumental mediation of role model effects on self-efficacy in politics. My studies attempted to expand on the initial attempt failing to detect such a link in survey data [19] using more direct measures and a 'two-fold' experimental design randomising both treatment and mediator, to improve causal inference.

I do, however, report evidence of instrumental links between descriptive representatives and voters. When women see women politicians in political groups, their judgement of (finance and healthcare) policy is more optimistic, the effect size being enough to close the gender gap in performance evaluations when politicians are only men. This is an important finding itself as it may have important implications for satisfaction with governance or with democratic institutions in general. Furthermore, improved policy evaluations may link to indicators of political engagement other than self-efficacy: external political efficacy (where the link is almost definitional, rather than causal), or trust (where evidence exist of causality [186]).

Group level competence versus self-efficacy?

Both studies suggest that confidence about political knowledge, an important part of survey inventories measuring internal political efficacy, is if anything *inversely related* to how much efficacy is attributed to politicians in executing policy. Looking at women, the effect size of attributing higher competence to groups is small but consistent across my

studies (a quarter confidence point differential higher versus lower picture rating in Study 1, control group versus mediator group membership in Study 2).

This may be indicative of a trade-off between self-efficacy and group competence, regardless of gender. In Study 1, effect sizes were comparable between men and women. Perhaps due to its specificity about woman friendly policy, Study 2 found large posterior variance for women while the mediator's effect size was more robust for men. As the current literature used in this work does not suggest hypotheses about this mechanism, future research might take it up to theorise how voters relate to elites when comparing policy competence: my initial evidence suggests politicians' competence judgements will peak when one's own competence is evaluated as low. The direction of causality here is unclear.

Outgroup effects

Study 2 suggests that, perhaps surprisingly, the instrumental mediation hypothesis stands if looking at men, not at women. Chapter 4 found evidence that men feel less confident about their own political knowledge and less efficacious when there are more women in visible political office. The current study suggests that some of these movements may be due to men's expectations about women empowering fellow women through substantive representation. This thinking is instrumental in that it disappeared when men were provided theoretical arguments that this may not be the case.

Notes

[29]There is no substantive reason why the link function is probit, rather than logit. In the R package used to fit Bayesian multilevel models, `MCMCglmm`[187], binomial logistic and ordinal probit options are readily available.

[30]Without prior specification, my estimates of effect sizes including their credible intervals are identical except an unrealistically large intercept making predicted probabilities very high in all categories (i.e. >99), making groups indistinguishable despite marginal effects.

[31]In the lab experiment, Chapter 5 Study 3, there was a reading task to complement emotion induction, about everyday sexism specifically.

Chapter 7

Conclusions

My thesis is motivated by questions about the role model effect: does an increasing presence of female politicians inspire women in the electorate to engage more with politics? What are the underlying mechanisms explaining why role models matter on the individual level? While empirical evidence exists answering the former question affirmatively, (a) it is predominantly observational, with uncertain claims of causality; (b) it often fails to explore a wide range of aspects of political engagement systematically, but reports isolated findings instead, such as a positive impact on political discussion; (c) if experimental, it is often reported in isolation from the insights provided by the more canonical public opinion surveys, with uncertain implications for electoral behaviour. My first empirical chapter thus sets out to contribute to the literature on descriptive representation by adding to the body of empirical evidence of role model effects and, uniquely, by bringing together experimental and observational approaches to measurement, reporting effects on a number of comparable dependent variables, with an emphasis on consistency across different approaches.

Turning to the second question, of why women in politics may serve as powerful role models when it comes to voter psychology, I first develop two sets of hypotheses: firstly, that the presence of women *alleviates stereotype threats* that harm women's domain identification, evidenced by state- and situation-specific anxieties; and secondly, that the presence of women in politics *improves public expectations about policy outcomes*, evidenced by competence judgements on the group level, and by specific expectations of women-friendly policies. Two empirical chapters test these hypotheses, using experimental methods that either allowed observation of the mediator variable's presence after randomised treatment delivery, or allowed randomisation of both treatment and mediator variables to improve causal inference.

My contributions here come against a background in previous literature in which first attempts tackling mediation were purely theoretical but untested, such as the role of self-interest and cost-benefit calculations [22],[30], while rare empirical contributions have failed propose a 'full' account of the causal chain, such as pointing at role of political discussion as a mediator role model effects, leaving further links between the presence of female politicians and the frequency of political discussion unexplored [19].

More detailed conclusions of the individual studies are available at the end of each section, as well as at the end of each chapter. In Chapter 4, Section 4.1.3 describes the results of an eye-tracking study, investigating attentional bias to gender balance in manipulated picture stimuli across a series of picture pairs, identical in details except that one was a 'contrast image' featuring predominantly or exclusively male politicians, and the other a 'role model image', always featuring an improved numerical balance. Section 4.2.5 draws joint conclusions from an investigation of psychological engagement with politics from a survey experiment and the British Election Study, where treatment or key predictor was exposure to manipulated gender balance in picture stimuli, or campaign exposure to gender balance among the 2010 and 2015 general election candidates, respectively.

In Chapter 5, Sections 5.1.3 and 5.2.4 infer anxiety mediation from two different experimental designs, while Section 5.3.4 isolates the differential impact of anger under stereotyping. In Chapter 6, Section 6.1.3 summarises the findings of my study exploring whether women's policy competence judgements of a political group improved when viewing an increasing proportion of women in them, across the issue areas of healthcare and finance. Section 6.2.3. summarises the findings of my investigation whether women's psychological engagement followed their expectations about substantive representation by female politicians. In the following sections, I assess evidence in support of the role model effect and of mediation, as well as interpret my results in terms of their broader theoretical and policy implications, concluding with suggestions of future studies that may address unexplored areas of my research.

Role model effects

Reviewing observational and experimental evidence, I concluded that attention and related variables were sensitive to 'tokenism', while a more linear role model effect described variables related to self-efficacy. Women's attention in the lab, or election interest and learning in the field, peaked under women's striking minority presence in campaign pictures or real-life campaigns; whereas female candidates failed to draw comparable attention if present in increasing proportions. By contrast, women became more and more confident in their own political awareness and voice in the 2015 constituencies where an increasing proportion of women ran, and became more and more confident in their political knowledge test answers in the survey experiment when viewing an increasing proportion of women in the campaign pictures.

My results thus point to a *plurality*, rather than a singularity, of role model effects thus far not reported in the empirical literature. A number of previous studies look at variation in candidate gender across time and space within American politics, where, often unintentionally, the only possible effect is that of tokenism. For example, studies looking at the gender composition of major party races are bound to observe only two candidates running in all male, mixed gender, or all female races [24], but latter at most time points is unobserved thus role model effects are in reality token effects. Another study counting

the number 'viable' women candidates in House, Senate, and governor elections admits that the observed range of their variable is merely 0 to 2, rather than six or above[32]. If very few races featured 2 viable women (the authors don't report this descriptively), their predictions are driven by token races as a simple measurement artefact[33][19]. By contrast, comparative studies such as those looking at the proportion of women in national legislatures evidently use linear predictors [26],[18], which are unable to detect tokenism, any other discrete effect, or even curvilinearity[34].

Importantly, my findings of tokenism are consistent with recent studies that are more explicit about women's *novelty* as a condition or moderating factor behind role model effects. One study finds that candidates needed to be new, challenging incumbents, and electorally viable, with a significant share of the popular vote (over 45%), to act as role models for young girls [21]. Another recent study attempts to isolate the novelty factor (being a new face in the race) from sheer 'presence' (category including also returning candidates) with weak or no evidence behind novelty but 'presence' still meant tokenism as a proportion of all candidates [188]. Looking at US House races, but this time at the discrete categories of men only, mixed-gender, and women only races, another recent study finds that only mixed-gender races increased political interest and decreased 'indifference' to candidates, but not women only races [129]. Across all three investigations, key effects impacted political interest, following the news, and the frequency of political discussion: all consistent with the findings in my thesis suggesting token spike of attention and its likely consequences.

Admittedly, the message of tokenism is not the most normatively appealing one. In motivating my research, I drew on theories praising women's greater presence in politics for its empowering effect. Tokenism, by contrast, may imply that rather than acting on a substantive mechanism of empowerment, women only drew attention as unique figures perceived to be 'out of place' or as political 'outsiders'. In the eye tracking study, I found preliminary evidence that men in striking minority position received a comparable attention surplus.

There are at least two reasons, however, why pessimism may not be warranted. Firstly, my linear role model effects positively impacted women's internal political efficacy and confidence, negatively impacted women's testing anxiety, and made women more optimistic about finance and healthcare policy outcomes, suggesting that there is still plenty of space for empowerment. Secondly, while there is greatest democratic potential in a politically interested and informed citizenry, also men fall short on this ideal despite masculine politics being 'descriptively' representative of them, suggesting only limited potential of a straight link between descriptive representation and political interest. Men and women both respond to novelty: a recent surge in the electoral success of political outsiders, both parties and candidates, shows that large segments of the vote are driven by newness and being 'out of place', without necessarily impacting alternative dimensions of candidate evaluations such as competence or general likeability.

Mediation

One out of four studies showed that a situation-specific 'testing anxiety', rather than diffuse state anxiety, explained why exposure to role models increased women participants' confidence in their test answers. This is consistent with the definition of stereotype threat that is explicit about situational factors, increasingly finding that performance under pressure is particularly sensitive to the effects of stereotyping. In my survey experiment, the testing anxiety item was was more explicit about the upcoming political knowledge quiz, whereas the state anxiety item was not, suggesting that 'men only' politics is a high threat context if women are prompted to evaluate their own skills (or efficacy, as found in the BES), but this threat was gradually alleviated once more and more women were present in campaign imagery.

A small study in Chapter 5 looked at unexplored affective dimensions of stereotype threat, finding that, besides anxiety, another likely emotion describing experience under stereotype threat are sadness and the feeling of helplessness. In these cases, women increasingly relied on role models, and avoided masculine politics. I isolated a different impact of anger, however, which itself motivated participants to confront the source of threat, 'men only' politics, without needing to rely on external actors.

My experiments looking at 'instrumental mediation' suggest that an explicit consideration of women's success in policy is not explanatory of role model effects. Women did become more optimistic about policy outcomes, and so did men at least when looking at issue areas where women are stereotypically seen as 'competent', but attributing better competence to politicians seems to have, if anything, decreased women's own confidence about political knowledge. This may be in line with literature on self-deflating social comparisons [69] documenting evidence that women make a series of strategic comparisons when exposed to the success of women as elites, importantly those who have high self-efficacy in leadership themselves would deny that they would draw inspiration from role models. Following this logic, it makes sense to interpret my results as those having higher political self-efficacy giving lower competency ratings to external actors.

Outgroup effects

Throughout my studies, I detected interesting outgroup effects. On the one hand, men's survey answers served as an important framework of comparison as the role model effect is about empowering women, relative to men, suggesting that a uniform effect on both genders may not be a case of a role model effect. On the other hand, my analyses showed patterns not hypothesised by the role model literature, by which null effects on men's political engagement would be sufficient against positive effects on women's political engagement. Instead, I detected *negative* role model effects on men's confidence and testing anxiety, suggesting sensitivity to losing a dominant position in politics or 'gender resentment'. Before these effects emerged under women's greater presence at critical mass and parity, however, I found minimal though consistent evidence of tokenism, suggesting that men first experienced a *stereotype boost* once reminded that women are only

present in politics in a small minority—this may have strengthened dominance orientations. Latter finding is consistent with literature engaging with the possibility that if group members who are advantaged by stereotypes receive reminders of this, may perform better [157]. Finally, Chapter 6 Study 2 found that expectations about women implementing women-friendly policies explained why men's confidence decreased, suggesting that if not women themselves, men do operate on this assumption, with consequences on political engagement.

Further contributions

My studies reported in Chapter 5 have implications to the stereotype threat scholarship itself, which among many other theories in social psychology faced criticism in the recent 'replication crisis' of experimental work. The early studies using the stereotype threat paradigm, cited in Chapter 2, are underpowered on small student samples. The empirical literature on stereotype threats is prone to publication bias which makes it very difficult to assess replicability [189]. One way the theory is able to defend itself is by showing a wide range of applications across domains, types of measures and types of samples, such as the studies I reviewed in my theoretical framework. My studies show, for the first time, that the stereotype threat framework works on women's political engagement. While political figures such as Barack Obama or Hillary Clinton have been hypothesised to alter perceptions of stereotype threat, the dependent variables were outside of the realm of politics. Chapter 5 Study 1 makes and explicit link between women's numerical balance in political office, and confidence about political knowledge, replicated across two data collection rounds, and consistent with self-efficacy effects detected in the BES.

Limitations and further work

There are a number of limitations to highlight. The scope of my thesis is limited to psychological engagement with politics, already a large set of potential dependent variables. The theoretical and empirical tools I introduced are, however, applicable to political engagement in general including participation such as voting, campaigning, contacting politicians, and not the least, ambition to run for office. Especially latter is an important avenue of future research to expand on—there is a very recent, unpublished study linking stereotype threat to ambition [190] yet with inconsistent effects—, with a recent significant contribution in research pointing out that the gap in political ambition opens up in adolescence [14], while perhaps not coincidentally recent role model papers report strongest subgroup effects on the younger segments as well [19],[21].

Another aspect I chose not to scrutinise is partisanship. While American parties are frequently included in inquiries of the role model effect with a hypothesis that co-partisan role models engage more [25], I was aiming at a more general argument about psychological engagement with politics. Partisans are likely to be more politically interested, and/or aware of their options and the candidates, which in turn may be explained by a host of other variables such as partisanship in social networks, the relevance of competing social

identities such as religion, values and programmatic congruence with parties (agreement on policies). An investigation of partisan 'moderators' might be novel in a comparative context looking at, for example, the relevance of incumbent or challenger support. In my chapter using the British Election Study, however, isolating partisans may risk losing the politically disengaged from my focus, which is exactly where role models are expected to be powerful.

Finally, another possibility of future research would be innovation in research design. I introduced the approach of combining observational and experimental evidence by comparison, and measuring similar variables looking for consistency, however, more formal linkages may facilitate this approach. Bayesian methods are increasingly used to combine insights from different studies, in which experimental findings may serve as priors, a formal hypothesis to be tested in observational studies. Another route is improving causal inference in observational studies by using quasi-experiments. The introduction of quota policies is a potential 'treatment' assignment, although rarely randomised like the Indian experience reported in Chapter 2 [61]. Redistricting has been used in the United States to show if minority turnout is boosted once, for instance, Latino seats are created, finding modest impacts [191],[192]. I used the BES panel design to control for pre-campaign interest, but panel designs may have further potential to study role model effects. Long-running panel studies may capture a small but significant portion of the citizenry moving across constituencies with different experiences, either historic or recent, with political role models. Exploiting such designs may give further evidence to the role model hypothesis while reserving very high external validity. Furthermore, the time component may allow further observation of role model effects embedded within contexts or shaped by changing gender roles.

Notes

[32]The theoretical maximum is above six if the three offices are contested by all female major party candidates (2) plus any independents or third party candidates.

[33]Rather than treating the variable, which takes values of either 0, 1, or 2, as a discrete or ordinal variable, the authors use it as a continuous predictor. A few number of 2s in the data set will at best moderate the effect of the 1s, insensitive to differential impacts.

[34]Theoretically it would be possible to tweak the functional form, such as including a square term, to test for a specific type of curvilinearity. Discrete spikes are, however, virtually impossible to model this way.

Appendix 1—Survey experiment questionnaire

Content starts on new page.

x2---In the first section, you'll be looking at five campaign pictures, all from the UK. Under the pictures, you will find an option to rate how well the politicians as a group would possibly manage given issues. The more stars you give the better job you think they would do. You are welcome to rely on your gut feelings or whatever impressions you may extract from the imagery. <mark>OPTIONS RANDOMISED:</mark>

- Healthcare.(1)
- Finance.(2)

Thank you. The next sections cover a mixture of political topics, including a test of political knowledge.

<mark>BLOCK STARTING HERE ASKED ONLY IN SECOND ROUND OF DATA COLLECTION.</mark>

<mark>a1 OR a2 RANDOMISED.</mark>

a1. Would you say you are experiencing any these emotions right now?

	No	Yes
Pleasant (1)	○	○
Relaxed (12)	○	○
Calm (4)	○	○
Uncomfortable (5)	○	○
Uneasy (6)	○	○
Worried (7)	○	○
Nervous (8)	○	○
Jittery (9)	○	○
Indecisive (10)	○	○
Anxious (20)	○	○
Uncertain (21)	○	○

a2 Regarding the test of political knowledge, would you say you are experiencing any these emotions right now?

	No	Yes
Pleasant (1)	○	○
Relaxed (12)	○	○
Calm (4)	○	○
Uncomfortable (5)	○	○
Uneasy (6)	○	○
Worried (7)	○	○
Nervous (8)	○	○
Jittery (9)	○	○
Indecisive (10)	○	○
Anxious (20)	○	○
Uncertain (21)	○	○

<mark>t1 & t1_text, t2 & t2_text, ctrl & ctrl_text RANDOMISED.</mark>

t1 Please take a moment to think about and briefly describe 3-5 things that do make you anxious. You don't need to take too long on this one. Few words or a sentence would suffice.

t1_text Please describe in more detail the one situation that makes you the most anxious so that someone reading will become anxious.   A few sentences would suffice here.

t2 Please take a moment to think about the following two findings taken from a recent expert study about women in politics in the UK. Once you're done, type 'done' in the text box below and click the arrow to continue. *"The notion of women's feminizing of politics is controversial [...] implying that as a group, they inherently share a set of policy preferences or ideology. Women are not a monolithic entity with a collective set of interests and beliefs. One could cite many examples of female political figures who represent the antithesis of what many regard as "women's interests" (e.g., Margaret Thatcher). Representation [...] is not a guarantee that women representatives will act for women. [...] In finance debates, findings show that the most significant [factors in] debate participation are relevant committee membership, Conservative party [membership], and seniority, holding other explanatory variables at constant. Sex was not significant."*

t2_text How would you describe the main finding of the authors? Should you need to read the text again, navigate back below.   A few sentences would suffice here.

ctrl Please take a moment to think about and briefly describe 3-5 grocery items or food you bought this week. You don't need to take too long on this one. Few words or a sentence would suffice.

ctrl_text Please describe in more detail one of those products so that someone reading will better know its properties (such packaging, colour, size etc.)   A few sentences would suffice here.

	Not interested at all (1)	Fairly uninterested (2)	Neither interested, nor uninterested (3)	Fairly interested (4)	Very interested (5)
How interested would you say you are in national politics?	○	○	○	○	○

a4/2 Please tell us if you think that the following statements are true or false. If you are not sure, use your best guess.

	On a scale from 1 to 10, how confident are you in this answer?										True (1)	False (2)
	Not confident at all (1)	2	3	4	5	6	7	8	9	Very confident (10)		
Members of the European Parliament are elected directly (1)	○	○	○	○	○	○	○	○	○	○	○	○
Not all members of Cabinet are MPs (2)	○	○	○	○	○	○	○	○	○	○	○	○
Registering to vote is optional (3)	○	○	○	○	○	○	○	○	○	○	○	○
The European Union consists of 12 member states (4)	○	○	○	○	○	○	○	○	○	○	○	○
The House of Commons has more power than the House of Lords (5)	○	○	○	○	○	○	○	○	○	○	○	○
There has to be a general election every 4 years (6)	○	○	○	○	○	○	○	○	○	○	○	○
You can only vote in your local election if you pay council tax (7)	○	○	○	○	○	○	○	○	○	○	○	○
Local councils have the power to set the school leaving age in their own area (8)	○	○	○	○	○	○	○	○	○	○	○	○
The Prime Minister is not an MP (9)	○	○	○	○	○	○	○	○	○	○	○	○
Every problem a constituent takes to an MP is debated in the Commons (10)	○	○	○	○	○	○	○	○	○	○	○	○
Councils can set the school leaving age (11)	○	○	○	○	○	○	○	○	○	○	○	○
You have to be on the electoral register to vote in the UK (12)	○	○	○	○	○	○	○	○	○	○	○	○

a4/2 Please match the following people to their jobs. Please note that this refers to Theresa May's current government since 2016 summer, discarding any impact the Snap Elections may bring in after 8 July 2017.

	On a scale from 1 to 10, how confident are you in this answer?										Pick answer (RANDOMISED order)												
	Not confident at all (1)	2	3	4	5	6	7	8	9	Very confident (10)	Philip Hammond (1)	Boris Johnson (2)	Jeremy Corbyn (3)	Amber Rudd (4)	John Bercow (5)	Catherine Ashton (6)	Vladimir Putin (7)	Angela Merkel (8)	Paolo Gentiloni (9)	Matteo Renzi (10)	Bashar al-Assad (11)	Jean-Cl Juncker (12)	None listed (13)
Chancellor of the Exchequer (1)	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Deputy Prime Minister (2)	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Leader of the Labour Party (3)	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Home secretary (4)	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Speaker of the HoC (5)	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
President of Russia (6)	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Chancellor of Germany (7)	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Prime minister of Italy (8)	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
President of Syria	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○

(9) President of the EU (10)	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○

a5 To what extent do you agree with the following statements?

	Strongly disagree (1)	Rather disagree (2)	Neither agree nor disagree (3)	Rather agree (4)	Strongly agree (5)
I understand the important political issues facing our country. (1)	○	○	○	○	○
It takes too much time and effort to be active in politics. (2)	○	○	○	○	○
It is often difficult for me to understand what happens in government and politics. (3)	○	○	○	○	○
Politicians don't care what people like me think. (4)	○	○	○	○	○

a6 How much do you trust the following people?

	A great deal (5)	A lot (4)	A moderate amount (3)	A little (2)	Absolutely no trust at all (1)
Members of Parliament in general? (1)	○	○	○	○	○
The MP of your local constituency? (2)	○	○	○	○	○
The UK police? (3)	○	○	○	○	○
People in general? (4)	○	○	○	○	○

2ND ROUND ONLY. a8 How important is your gender to your sense of who you are?_____ (1)

2ND ROUND ONLY. a9 How much do you agree with the following two statements?

_____ In general, things would improve if there were more women in politics. (1)
_____ Women should be given preferential treatment when applying for jobs and promotions. (2)

a10 How much do you agree with the following two statements?

	Strongly disagree (1)	Rather disagree (2)	Neither agree nor disagree (3)	Rather agree (4)	Strongly agree (5)
If certain groups stayed in their place, we would have fewer problems (1)	○	○	○	○	○
It's probably a good thing that certain groups are at the top and other groups are at the bottom (2)	○	○	○	○	○
Group equality should be our ideal (3)	○	○	○	○	○
We should do what we can to equalise conditions for everyone (4)	○	○	○	○	○

a11 Please indicate your attitude towards the following issues by giving thumbs up, thumbs down, or 'unsure'.

	Thumbs down 👎 (1)	Unsure ? (2)	Thumbs up 👍 (3)
Abortion (1)	○	○	○
Limited government (2)	○	○	○
Military and national security (3)	○	○	○
Religion (4)	○	○	○
Welfare benefits (5)	○	○	○
Gun ownership (6)	○	○	○
Traditional marriage (7)	○	○	○
Traditional values (8)	○	○	○
Fiscal responsibility (9)	○	○	○
Business (10)	○	○	○
The family unit (11)	○	○	○
Britishness (12)	○	○	○
Queen Elizabeth II. (13)	○	○	○
Brexit (14)	○	○	○

a12 What year were you born?

a13 What is the highest educational qualification you have?
- ○ Incomplete Secondary Education (below GCSE/ O-Levels) (1)
- ○ Secondary Education Completed (GCSE / O Level / CSE or equivalent) (2)
- ○ Secondary Education Completed (A Level or equivalent) (3)
- ○ Some vocational or Technical Qualifacations (4)
- ○ University Education Completed (First Degree, e.g. BA or BSc) (5)
- ○ Vocational or Technical Qualifications (e.g. HND or NVQ) (6)
- ○ Postgraduate Degree or equivalent (7)
- ○ Doctorate, Post-doctorate or equivalent (Higher Degree) (8)
- ○ Prefer not to answer ()

Appendix 2—Model diagnostics

The upcoming plots show convergence of both fixed- (in all tables labelled as Betas) and random effects (in all tables labelled as Gamma) in 33 selected models. In Chapter 4, my analytical strategy was to first fit uninformative or 'null' models, later expanded to include gender, treatment, and all covariates; in these cases, I only generated a convergence plot for the largest extended model.

Convergence itself is necessary to interpret point estimates: if the trace plots show a trend such as an increase or a decrease, the reported posterior means (the point estimates in my model tables) are not (yet) reliable. For example, a mean β of 'Gender', estimated at 1.00 but calculated from a posterior sample where trace plot shows an increasing trend, is likely to be converge later at a higher value.

There are two general remarks to make. Fixed-effects convergence is acceptable in all models except in some of my generalised linear models, such as my poisson models 3.2, and 5.2. This was part of my reasoning not to continue fitting models on saccadic data in Chapter 5 but concentrate on fixation duration. Nevertheless, in extensions of this study I may experiment with longer Markov Chains to see if convergence is satisfactory, before concluding on null results. Another option would be to experiment with informative priors, on a case by case basis. This approach may belong in shorter reports, as here I aimed at uniform solutions to increase comparability across models.

Random effects convergence is very often unsatisfactory. This is a heavily debated issue in the multilevel modelling literature: with only a few groups such as data collection rounds, it is increasingly difficult to estimate variance components. At the problematic models, however, random effects convergence is not a major issue because they accompanied large, often 99% residual variances with an upward trend. In these cases I merely concluded that those variances are likely to be negligible, and I explicitly reported if I obtained models that have low explanatory powers. Nevertheless, in later explorations, I may need to experiment on a case-by-case basis to use informative priors that might fix this problem.
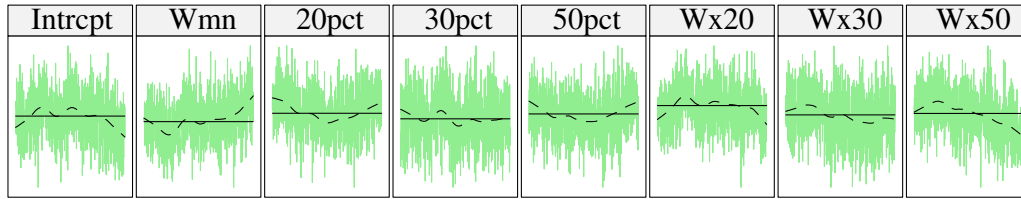
Content starts on new page.
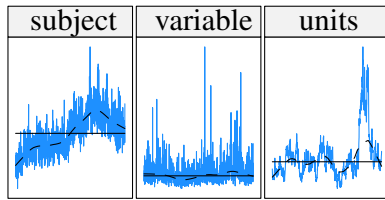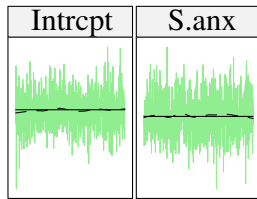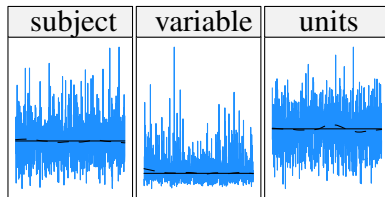
Convergence of fixed effects, Model 1.2.
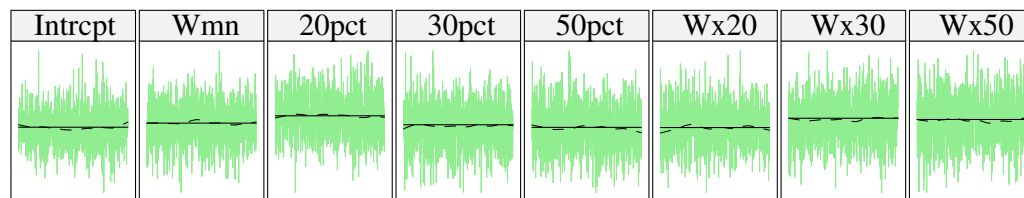
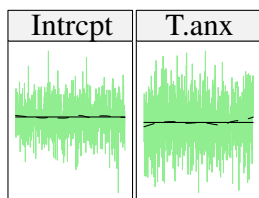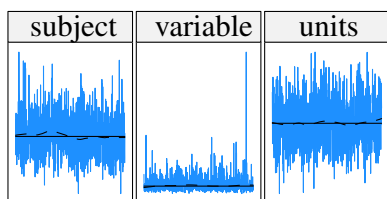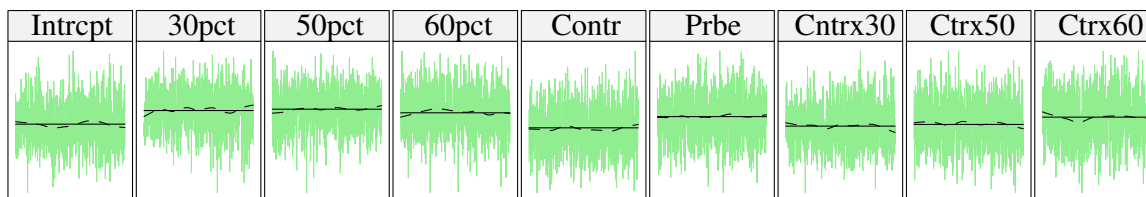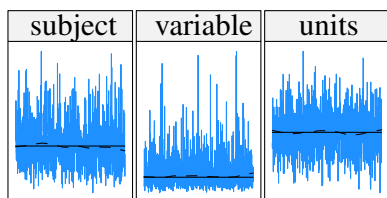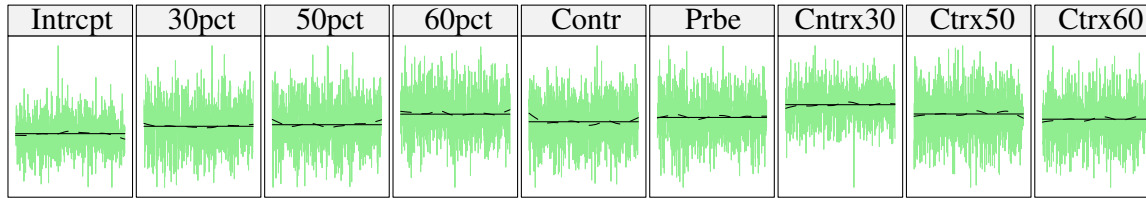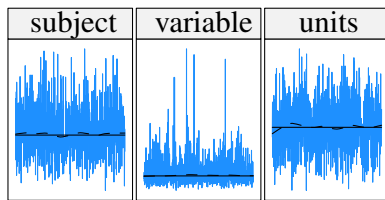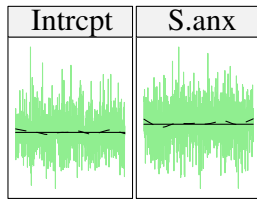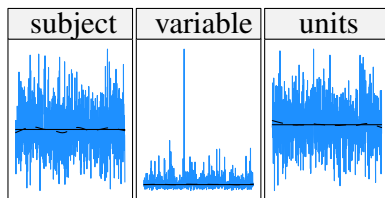*Convergence of random effects, Model 1.2.*
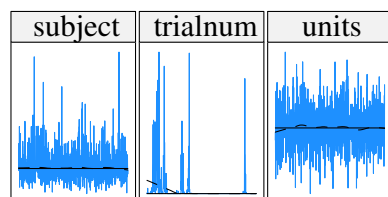
Convergence of fixed effects, Model 2.2.
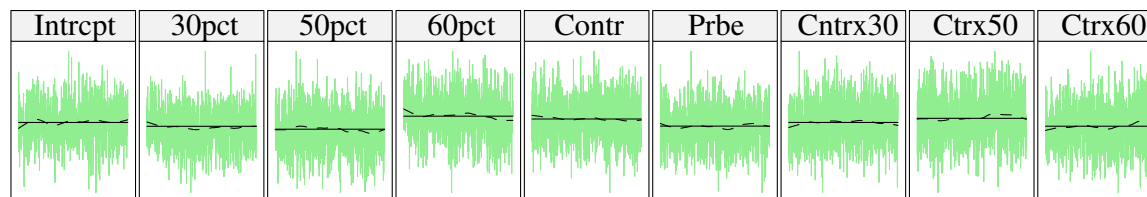
Convergence of random effects, Model 2.2.

Convergence of fixed effects, Model 3.2.

| Intrcpt | 30pct | 50pct | 60pct | Contr | Prbe | Cntrx30 | Ctrx50 | Ctrx60 |

Convergence of random effects, Model 3.2.

| subject | trial | units |

Convergence of fixed effects, Model 4.2.

| Intrcpt | 30pct | 50pct | 60pct | Contr | Prbe | Cntrx30 | Ctrx50 | Ctrx60 |
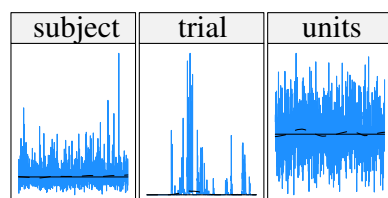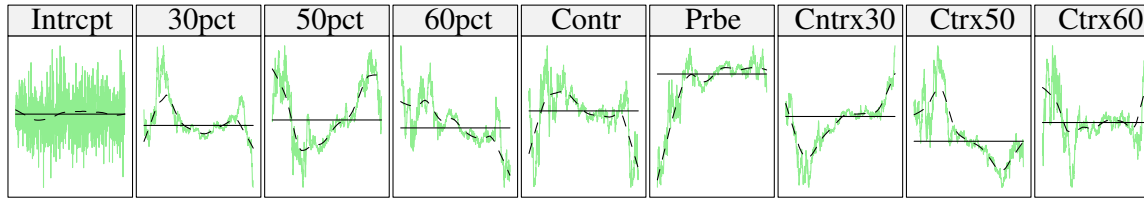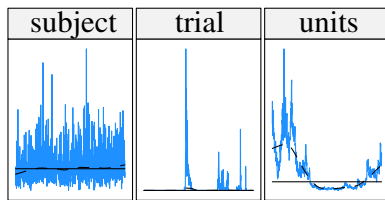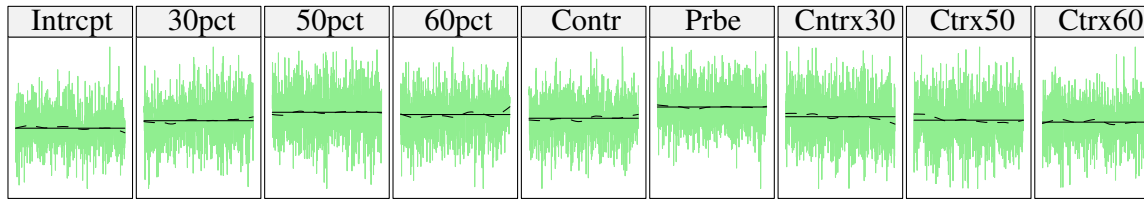
Convergence of random effects, Model 4.2.
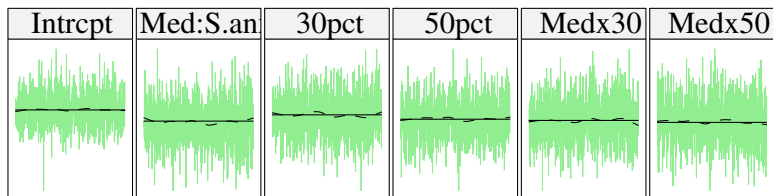
| subject | trial | units |
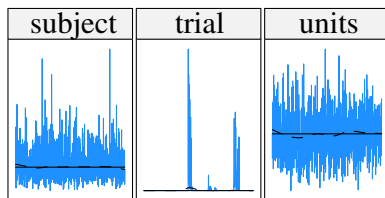
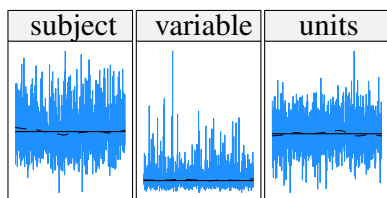*Convergence of fixed effects, Model 5.2.*
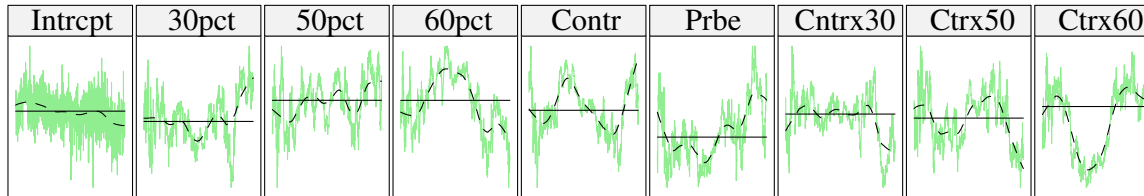
Convergence of random effects, Model 5.2.
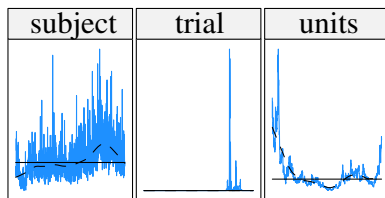
Convergence of fixed effects, Model 6.2.
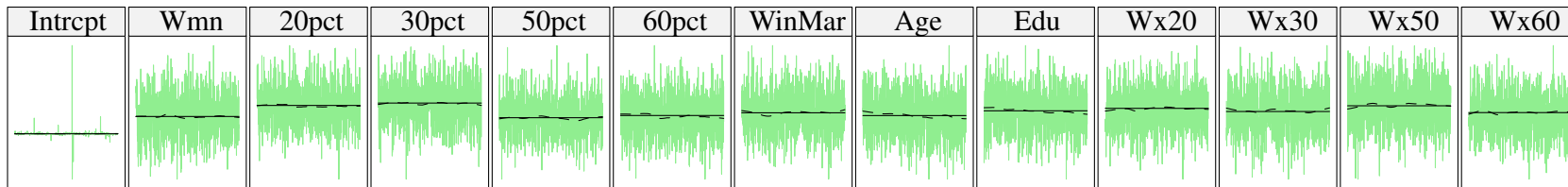
Convergence of random effects, Model 6.2.
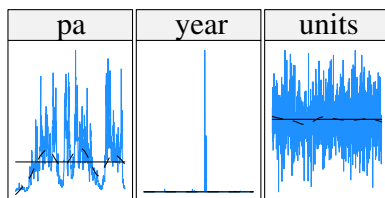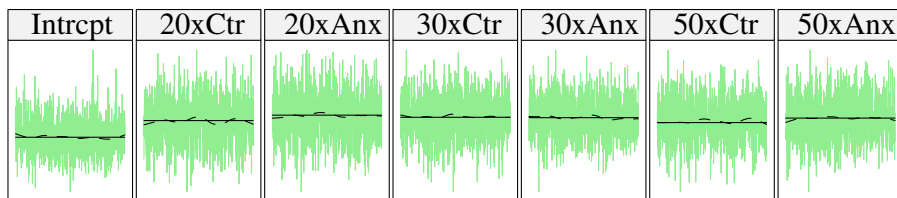
Convergence of fixed effects, Model 7.2.
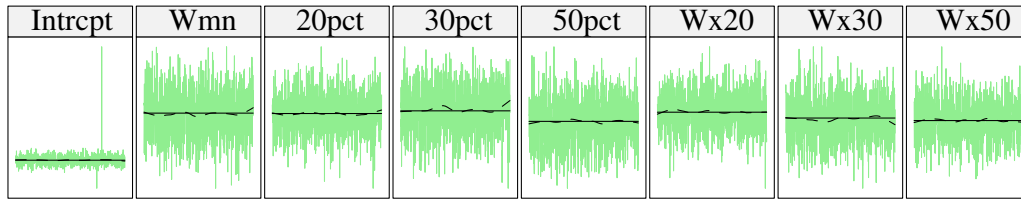
Convergence of random effects, Model 7.2.
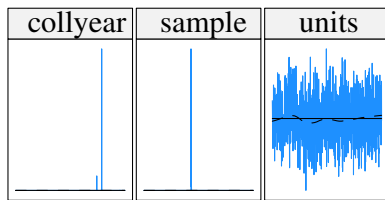
Convergence of fixed effects, Model 8.2.
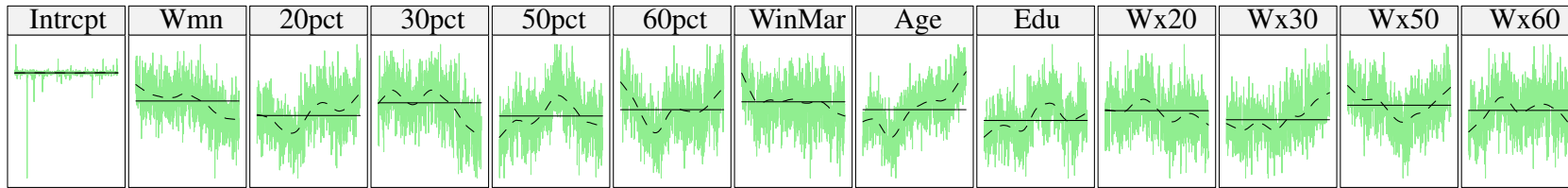
Convergence of random effects, Model 8.2.
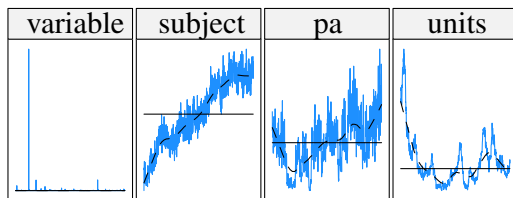
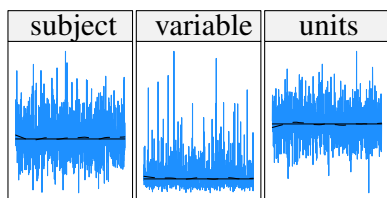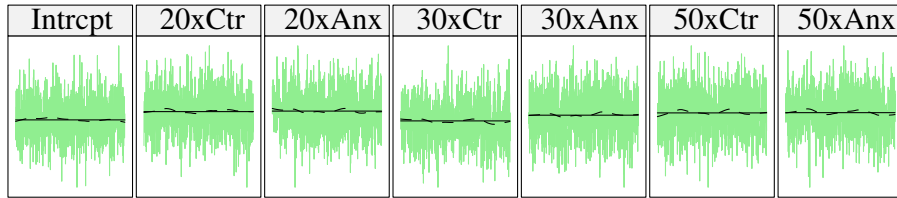*Convergence of fixed effects, Model 9.2.*
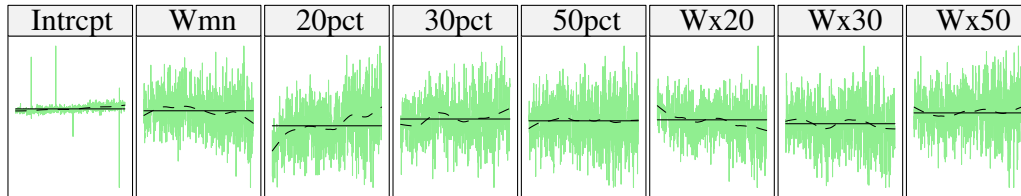
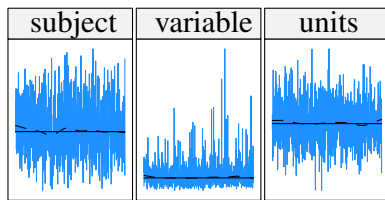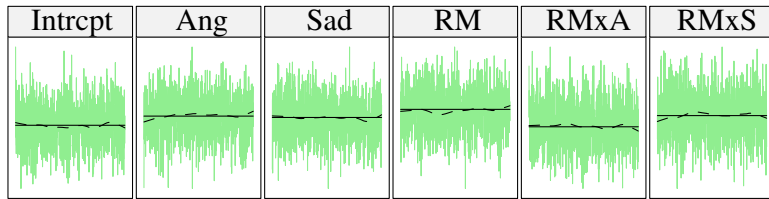*Convergence of random effects, Model 9.2.*
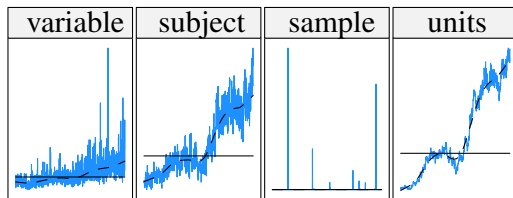
Convergence of fixed effects, Model 10.2.
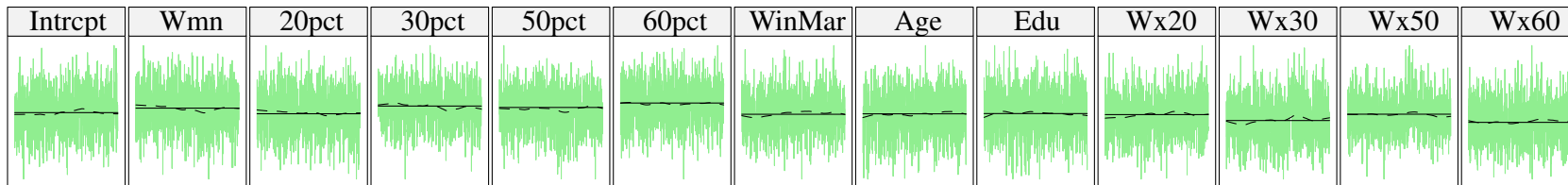
Convergence of random effects, Model 10.2.
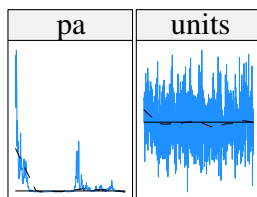
Convergence of fixed effects, Model 11.2.
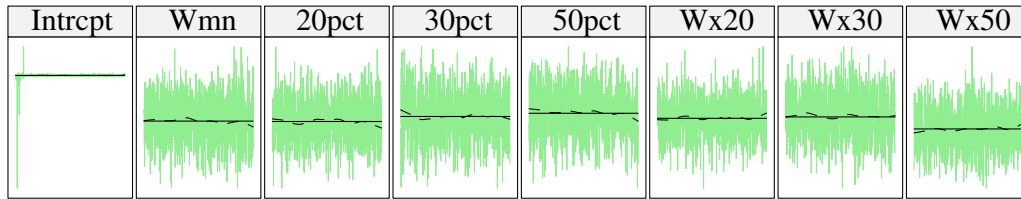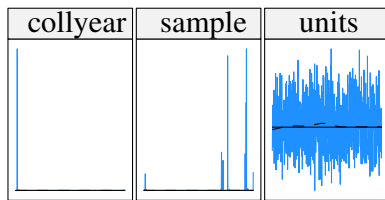
Intrcpt	Wmn	20pct	30pct	50pct	Wx20	Wx30	Wx50

Convergence of random effects, Model 11.2.

collyear	sample	units

Convergence of fixed effects, Model 12.2.
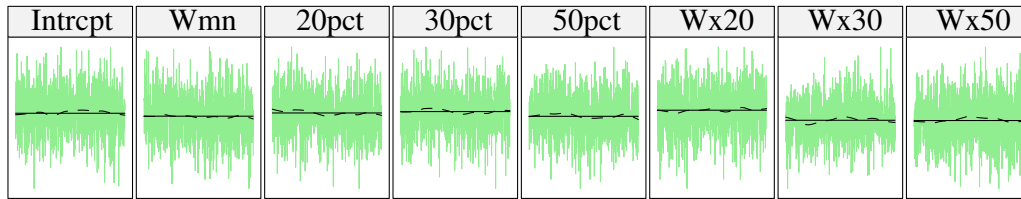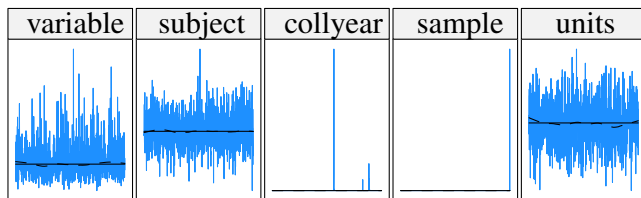
Intrcpt	Wmn	20pct	30pct	50pct	Wx20	Wx30	Wx50

Convergence of random effects, Model 12.2.

variable	subject	collyear	sample	units

Convergence of fixed effects, Model 13.2.

| Intrcpt | Wmn | 20pct | 30pct | 50pct | 60pct | WinMar | Age | Edu | Wx20 | Wx30 | Wx50 | Wx60 |

Convergence of random effects, Model 13.2.

| pa | units |

Convergence of fixed effects, Model 14.2.

| Intrcpt | Wmn | 20pct | 30pct | 50pct | 60pct | WinMar | Age | Edu | Wx20 | Wx30 | Wx50 | Wx60 |

Convergence of random effects, Model 14.2.

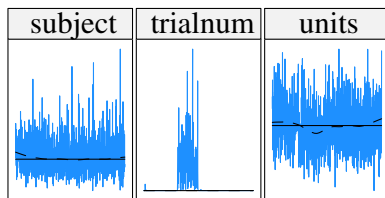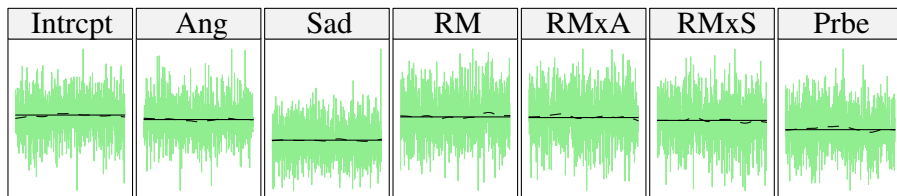| variable | subject | pa | type | units |
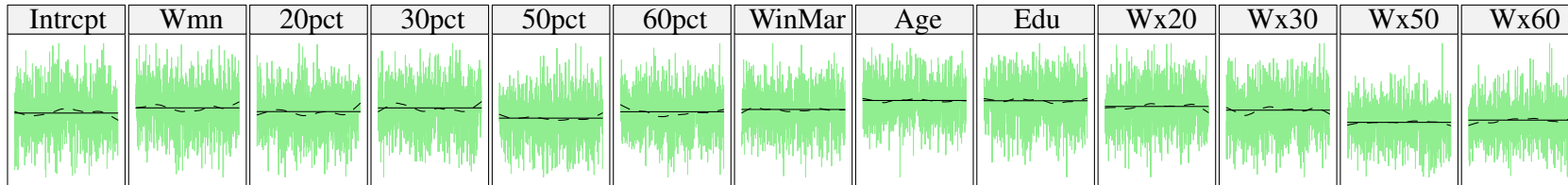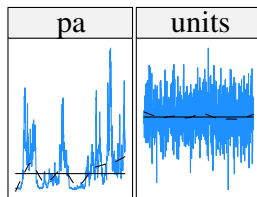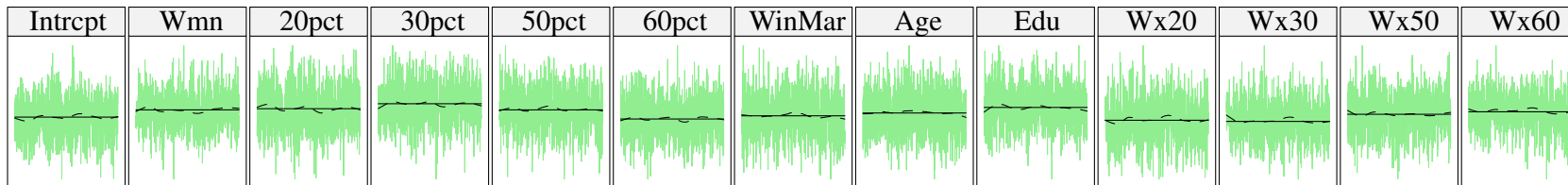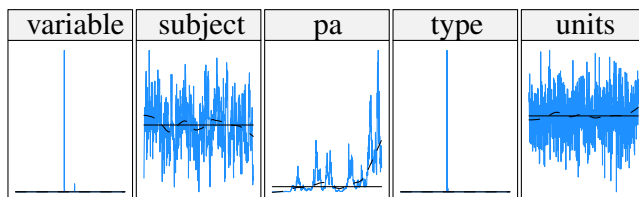
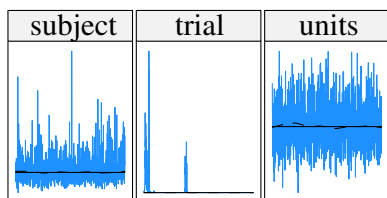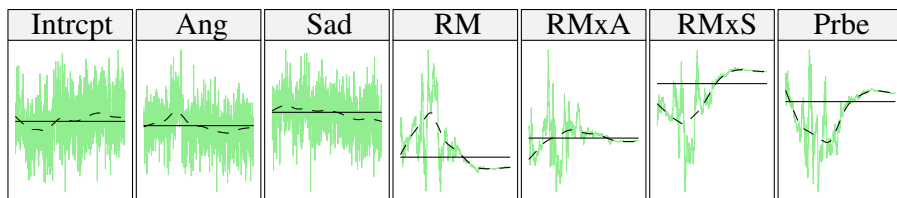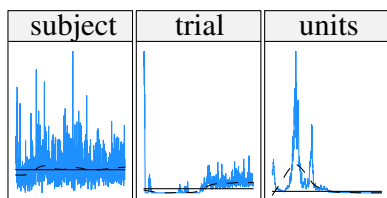*Convergence of fixed effects, Model 15.2.*
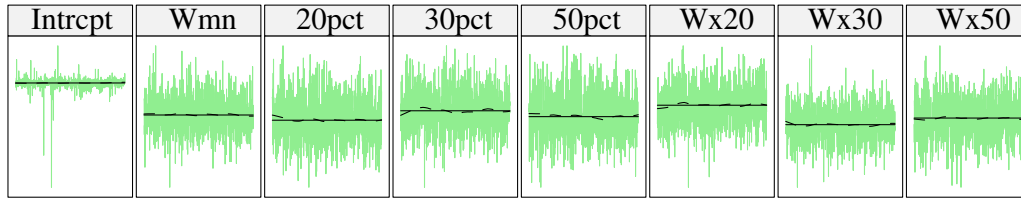
Convergence of random effects, Model 15.2.
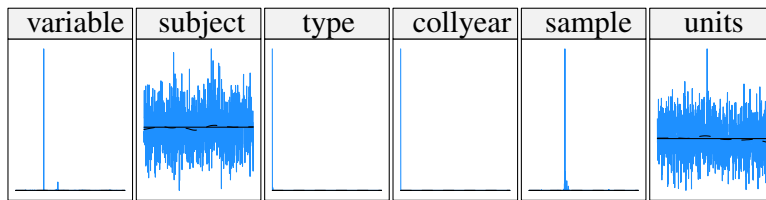
Convergence of fixed effects, Model 16.2.
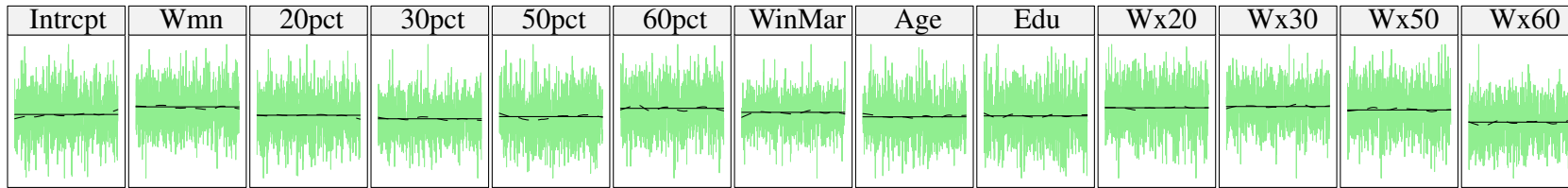
Convergence of random effects, Model 16.2.
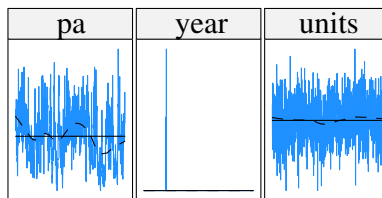
Convergence of fixed effects, Model 17.2.
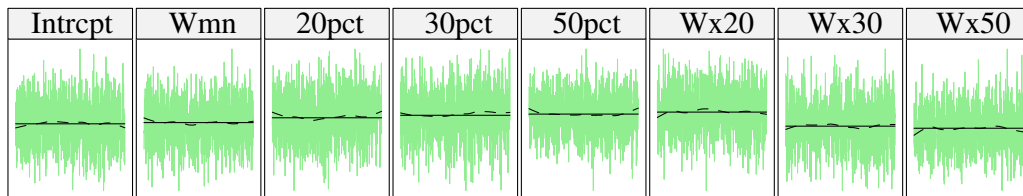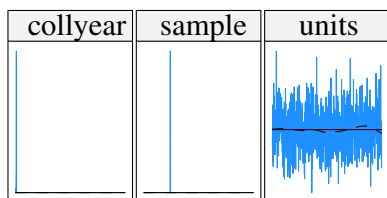
Intrcpt	Wmn	20pct	30pct	50pct	Wx20	Wx30	Wx50

Convergence of random effects, Model 17.2.

collyear	sample	units

Convergence of fixed effects, Model 18.
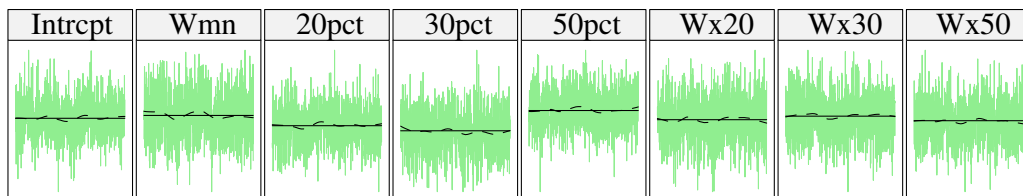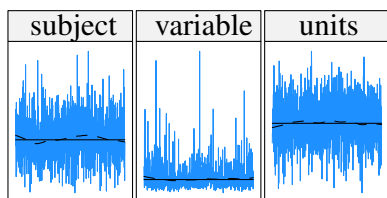
Intrcpt	Wmn	20pct	30pct	50pct	Wx20	Wx30	Wx50

Convergence of random effects, Model 18.

subject	variable	units

Convergence of fixed effects, Model 19.2.

Intrcpt	Wmn	20pct	30pct	50pct	Wx20	Wx30	Wx50

Convergence of random effects, Model 19.2.

subject	variable	units

Convergence of fixed effects, Model 20.1.

Intrcpt	S.anx

Convergence of random effects, Model 20.1.

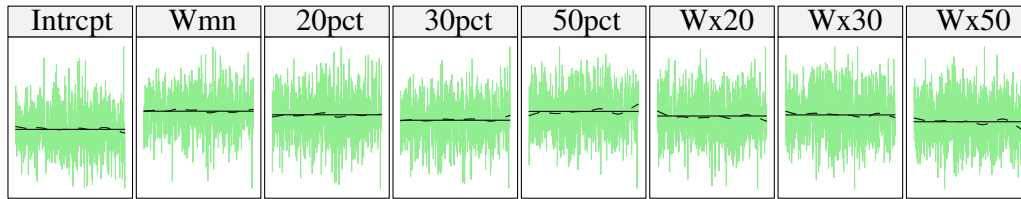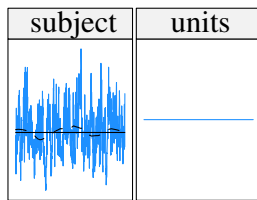| subject	variable	units

Convergence of fixed effects, Model 20.2.

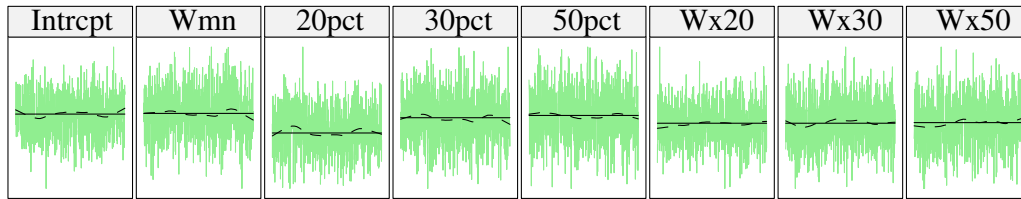*Convergence of random effects, Model 20.2.*
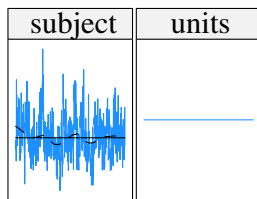

Convergence of fixed effects, Model 21.1.

Convergence of random effects, Model 21.1.
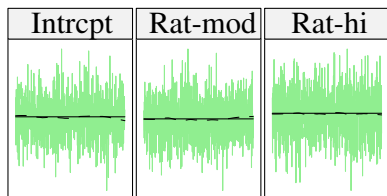
Convergence of fixed effects, Model 21.2.

Convergence of random effects, Model 21.2.

Convergence of fixed effects, Model 22.

Convergence of random effects, Model 22.

Convergence of fixed effects, Model 23.

Intrcpt	Med:S.an	30pct	50pct	Medx30	Medx50

Convergence of random effects, Model 23.

subject	variable	units

Convergence of fixed effects, Model 24.1.

Intrcpt	20xCtr	20xAnx	30xCtr	30xAnx	50xCtr	50xAnx

Convergence of random effects, Model 24.1.

subject	variable	units

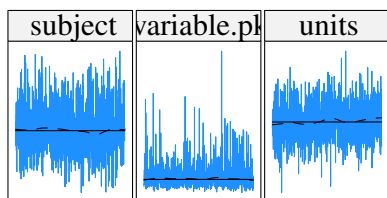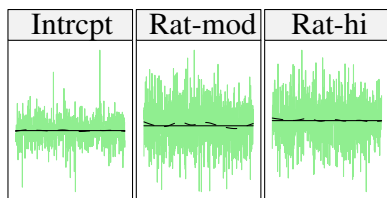*Convergence of fixed effects, Model 24.2.*

Convergence of random effects, Model 24.2.

Convergence of fixed effects, Model 25.

Convergence of random effects, Model 25.

Convergence of fixed effects, Model 26.

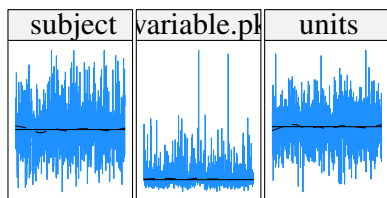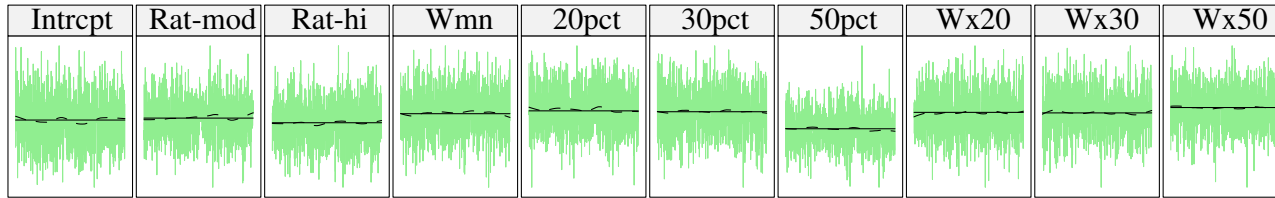*Convergence of random effects, Model 26.*

Convergence of fixed effects, Model 27.

Convergence of random effects, Model 27.

Convergence of fixed effects, Model 28.1.

Intrcpt	Wmn	20pct	30pct	50pct	Wx20	Wx30	Wx50

Convergence of random effects, Model 28.1.

subject	units

Convergence of fixed effects, Model 28.2.

Intrcpt	Wmn	20pct	30pct	50pct	Wx20	Wx30	Wx50

Convergence of random effects, Model 28.2.

subject	units

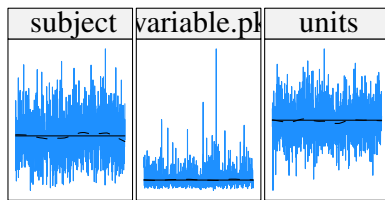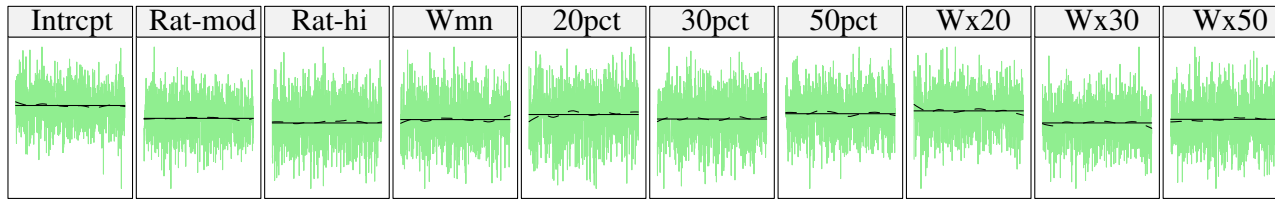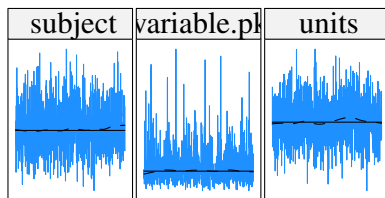*Convergence of fixed effects, Model 29.1.*

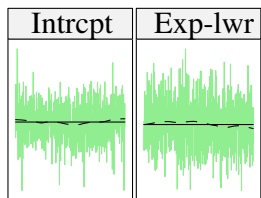*Convergence of random effects, Model 29.1.*

Convergence of fixed effects, Model 29.2.

Convergence of random effects, Model 29.2.

Convergence of fixed effects, Model 30.1.

Intrcpt	Rat-mod	Rat-hi	Wmn	20pct	30pct	50pct	Wx20	Wx30	Wx50

Convergence of random effects, Model 30.1.

subject	variable.pl	units

Convergence of fixed effects, Model 30.2.

Intrcpt	Rat-mod	Rat-hi	Wmn	20pct	30pct	50pct	Wx20	Wx30	Wx50

Convergence of random effects, Model 30.2.

subject	variable.pl	units

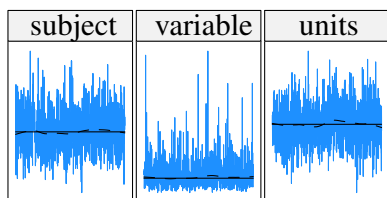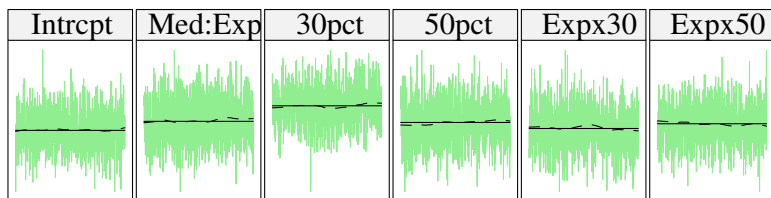*Convergence of fixed effects, Model 31.*

Convergence of random effects, Model 31.

Convergence of fixed effects, Model 32.

Convergence of random effects, Model 32.

Convergence of fixed effects, Model 33.1.

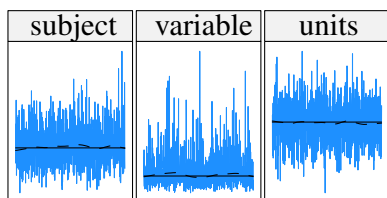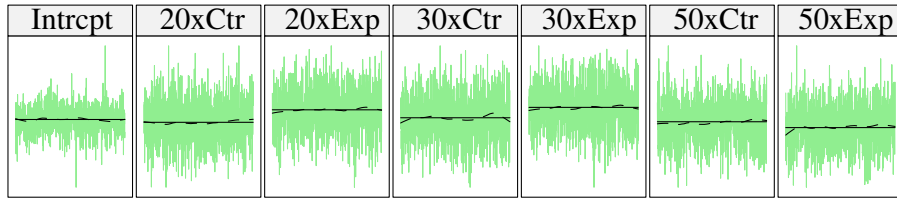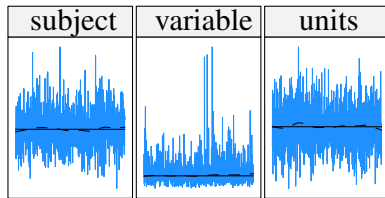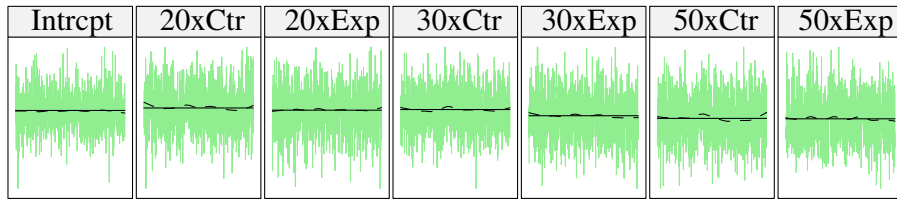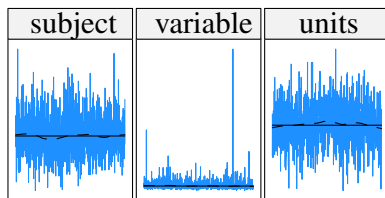*Convergence of random effects, Model 33.1.*

Convergence of fixed effects, Model 33.2.

Convergence of random effects, Model 33.2.

Bibliography

[1] @HillaryClinton on Twitter. Accessed at https://twitter.com/HillaryClinton/status/740349871073398785.

[2] S. Vedantam, "How stereotypes can drive women to quit science," *National Public Radio*, 2012. Accessed at https://www.npr.org/2012/07/12/156664337/stereotype-threat-why-women-quit-science-jobs.

[3] "Women on boards. HM Government Equalities Office Report," 2011. Accessed at https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/31480/11-745-women-on-boards.pdf.

[4] "Gender Differences in Voter Turnout," Fact Sheet, Center for American Women and Politics, 2015. Accessed at http://cawp.rutgers.edu/sites/default/files/resources/genderdiff.pdf.

[5] P. Norris, "Gender: A Gender-Generation Gap?," in *Critical Elections: British Parties and Voters in Long-term Perspective* (G. Evans and P. Norris, eds.), p. 153, 153: London: Sage, 1999.

[6] H. Coffé and C. Bolzendahl, "Same Game, Different Rules? Gender Differences in Political Participation," *Sex Roles*, vol. 62, no. 5-6, pp. 318–333, 2010.

[7] S. Verba, N. Burns, and K. L. Schlozman, "Knowing and Caring about Politics: Gender and Political Engagement," *The Journal of Politics*, vol. 59, no. 4, pp. 1051–1072, 1997.

[8] S. Verba, K. L. Schlozman, and H. E. Brady, *Voice and Equality: Civic Voluntarism in American Politics*. Harvard University Press, 1995.

[9] L. L. Bennett and S. E. Bennett, "Enduring Gender Differences in Political Interest: The impact of socialization and political dispositions," *American Politics Quarterly*, vol. 17, no. 1, pp. 105–122, 1989.

[10] M. X. Delli Carpini and S. Keeter, "The gender gap in political knowledge," *The Public Perspective*, vol. 3, no. 5, p. 23, 1992.

[11] J. Wolak and M. McDevitt, "The Roots of the Gender Gap in Political Knowledge in Adolescence," *Political Behavior*, vol. 33, no. 3, pp. 505–533, 2011.

[12] P. Norris, J. Lovenduski, and R. Campbell, "Gender and political participation," research report, The Electoral Commission, 2004. Accessed at: .

[13] J. L. Lawless and R. L. Fox, *It Still Takes A Candidate: Why Women Don't Run for Office*. Cambridge University Press, 2010.

[14] J. L. Lawless and R. L. Fox, *Running from Office: Why Young Americans are Turned Off to Politics*. Oxford University Press, USA, 2015.

[15] S. Verba, N. H. Nie, and J.-o. Kim, *Participation and Political Equality: A Seven-Nation Comparison*. University of Chicago Press, 1978.

[16] S. Welch, "Sex differences in political activity in Britain," *Women & Politics*, vol. 1, no. 2, pp. 29–46, 1980.

[17] N. Burns, K. L. Schlozman, and S. Verba, *The Private Roots of Public Action: Gender, Equality, and Political Participation*. Cambridge, Mass: Harvard University Press, 2001.

[18] C. Wolbrecht and D. E. Campbell, "Leading by example: Female members of parliament as political role models," *American Journal of Political Science*, vol. 51, no. 4, pp. 921–939, 2007.

[19] D. E. Campbell and C. Wolbrecht, "See Jane run: Women politicians as role models for adolescents," *The Journal of Politics*, vol. 68, no. 2, pp. 233–247, 2006.

[20] S.-J. S. Liu and L. A. Banaszak, "Do government positions held by women matter? A cross-national examination of female ministers' impacts on women's political participation," *Politics & Gender*, vol. 13, no. 1, pp. 132–162, 2017.

[21] C. Wolbrecht and D. E. Campbell, "Role models revisited: Youth, novelty, and the impact of female candidates," *Politics, Groups, and Identities*, vol. 5, pp. 1–17, 2017.

[22] L. Bobo and F. D. Gilliam, "Race, Sociopolitical Participation, and Black Empowerment," *American Political Science Review*, vol. 84, no. 2, pp. 377–393, 1990.

[23] S. B. Hansen, "Talking About Politics: Gender and Contextual Effects on Political Proselytizing," *The Journal of Politics*, vol. 59, no. 1, pp. 73–103, 1997.

[24] L. R. Atkeson, "Not All Cues Are Created Equal: The Conditional Impact of Female Candidates on Political Engagement," *The Journal of Politics*, vol. 65, no. 4, pp. 1040–1061, 2003.

[25] B. Reingold and J. Harrell, "The Impact of Descriptive Representation on Women's Political Engagement: Does Party Matter?," *Political Research Quarterly*, vol. 63, no. 2, pp. 280–294, 2010.

[26] J. A. Karp and S. A. Banducci, "When politics is not just a man's game: Women's representation and political engagement," *Electoral Studies*, vol. 27, no. 1, pp. 105 – 115, 2008.

[27] J. Wolak, "Candidate Gender and the Political Engagement of Women and Men," *American Politics Research*, vol. 43, no. 5, p. 872, 2015.

[28] J. L. Lawless, "Politics of Presence? Congresswomen and Symbolic Representation," *Political Research Quarterly*, vol. 57, no. 1, pp. 81–99, 2004.

[29] K. Dolan, "The impact of candidate sex on evaluations of candidates for the US House of Representatives," *Social Science Quarterly*, vol. 85, no. 1, pp. 206–217, 2004.

[30] J. Mansbridge, "Should Blacks Represent Blacks and Women Represent Women? A Contingent "Yes"," *The Journal of Politics*, vol. 61, no. 3, pp. 628–657, 1999.

[31] J. Sherman, "Automatic and controlled components of implicit stereotyping and prejudice," *Psychological Science Agenda*, vol. 19, no. 3, 2005.

[32] C. M. Steele and J. Aronson, "Stereotype threat and the intellectual test performance of African Americans," *Journal of Personality and Social Psychology*, vol. 69, no. 5, pp. 797–811, 1995.

[33] D. M. Amodio, E. Harmon-Jones, and P. G. Devine, "Individual differences in the activation and control of affective race bias as assessed by startle eyeblink response and self-report.," *Journal of Personality and Social Psychology*, vol. 84, no. 4, p. 738, 2003.

[34] S. J. Spencer, C. Logel, and P. G. Davies, "Stereotype threat," *Annual Review of Psychology*, vol. 67, pp. 415–437, 2016.

[35] J. Aronson, C. B. Fried, and C. Good, "Reducing the effects of stereotype threat on african american college students by shaping theories of intelligence," *Journal of Experimental Social Psychology*, vol. 38, no. 2, pp. 113–125, 2002.

[36] R. P. Brown and E. A. Day, "The difference isn't black and white: stereotype threat and the race gap on raven's advanced progressive matrices.," *Journal of Applied Psychology*, vol. 91, no. 4, p. 979, 2006.

[37] S. J. Spencer, C. M. Steele, and D. M. Quinn, "Stereotype Threat and Women's Math Performance," *Journal of Experimental Social Psychology*, vol. 35, no. 1, pp. 4–28, 1999.

[38] P. M. Gonzales, H. Blanton, and K. J. Williams, "The effects of stereotype threat and double-minority status on the test performance of latino women," *Personality and Social Psychology Bulletin*, vol. 28, no. 5, pp. 659–670, 2002.

[39] D. Abrams, A. Eller, and J. Bryant, "An age apart: the effects of intergenerational contact and stereotype threat on performance and intergroup bias.," *Psychology and Aging*, vol. 21, no. 4, pp. 691–702, 2006.

[40] H.-H. D. Nguyen and A. M. Ryan, "Does stereotype threat affect test performance of minorities and women? a meta-analysis of experimental evidence.," 2008.

[41] C. M. Steele, "A threat in the air: How stereotypes shape intellectual identity and performance.," *American Psychologist*, vol. 52, no. 6, p. 613, 1997.

[42] B. W. Griffin, "Academic disidentification, race, and high school dropouts," *The High School Journal*, vol. 85, no. 4, pp. 71–81, 2002.

[43] P. G. Davies, S. J. Spencer, D. M. Quinn, and R. Gerhardstein, "Consuming images: How television commercials that elicit stereotype threat can restrain women academically and professionally," *Personality and Social Psychology Bulletin*, vol. 28, no. 12, pp. 1615–1628, 2002.

[44] J. R. Shapiro and A. M. Williams, "The Role of Stereotype Threats in Undermining Girls' and Women's Performance and Interest in STEM Fields," *Sex Roles*, vol. 66, no. 3-4, pp. 175–183, 2012.

[45] M. C. Murphy, C. M. Steele, and J. J. Gross, "Signaling threat: How situational cues affect women in math, science, and engineering settings," *Psychological Science*, vol. 18, no. 10, pp. 879–885, 2007.

[46] S. Cheryan, S. A. Ziegler, A. K. Montoya, and L. Jiang, "Why are some STEM fields more gender balanced than others?," *Psychological Bulletin*, vol. 143, no. 1, p. 1, 2017.

[47] H. R. Milner and A. W. Hoy, "A case study of an african american teacher's self-efficacy, stereotype threat, and persistence," *Teaching and Teacher Education*, vol. 19, no. 2, pp. 263–276, 2003.

[48] F. Reichert, "How internal political efficacy translates political knowledge into political participation: Evidence from germany," *Europe's Journal of Psychology*, vol. 12, no. 2, p. 221, 2016.

[49] G. I. Balch, "Multiple indicators in survey research: The concept" sense of political efficacy"," *Political Methodology*, pp. 1–43, 1974.

[50] M. S. McGlone, J. Aronson, and D. Kobrynowicz, "Stereotype Threat and the Gender Gap in Political Knowledge," *Psychology of Women Quarterly*, vol. 30, no. 4, pp. 392–398, 2006.

[51] J. J. Mondak and M. R. Anderson, "The knowledge gap: A reexamination of gender-based differences in political knowledge," *The Journal of Politics*, vol. 66, no. 2, pp. 492–512, 2004.

[52] F. Batista Pereira, "Women's representation, stereotype threat, and the gender gap in political knowledge," 2014. Unpublished conference paper. Accessed at http://polisci.utk.edu/womenpolitics/docs/Batista%20-%20UTworkshop%20paper.pdf.

[53] S. Pruysers and J. Blais, "Why won't lola run? an experiment examining stereotype threat and political ambition," *Politics & Gender*, vol. 13, no. 2, pp. 232–252, 2017.

[54] D. M. Marx and J. S. Roman, "Female role models: Protecting women's math test performance," *Personality and Social Psychology Bulletin*, vol. 28, no. 9, pp. 1183–1193, 2002.

[55] R. B. McIntyre, R. M. Paulson, and C. G. Lord, "Alleviating women's mathematics stereotype threat through salience of group achievements," *Journal of Experimental Social Psychology*, vol. 39, no. 1, pp. 83–90, 2003.

[56] D. M. Marx, S. J. Ko, and R. A. Friedman, "The 'Obama effect': How a salient role model reduces race-based performance differences," *Journal of Experimental Social Psychology*, vol. 45, no. 4, pp. 953–956, 2009.

[57] J. Aronson, S. Jannone, M. McGlone, and T. Johnson-Campbell, "The Obama effect: An experimental test," *Journal of Experimental Social Psychology*, vol. 45, no. 4, pp. 957–960, 2009.

[58] C. A. Taylor, C. G. Lord, R. B. McIntyre, and R. M. Paulson, "The Hillary Clinton effect: When the same role model inspires or fails to inspire improved performance under stereotype threat," *Group Processes & Intergroup Relations*, vol. 14, no. 4, pp. 447–459, 2011.

[59] I. M. Latu, M. S. Mast, J. Lammers, and D. Bombari, "Successful female leaders empower women's behavior in leadership tasks," *Journal of Experimental Social Psychology*, vol. 49, no. 3, pp. 444–448, 2013.

[60] M. Mariani, B. W. Marshall, and A. L. Mathews-Schultz, "See Hillary Clinton, Nancy Pelosi, and Sarah Palin run? Party, ideology, and the influence of female role models on young women," *Political Research Quarterly*, vol. 68, no. 4, pp. 716–731, 2015.

[61] L. Beaman, E. Duflo, R. Pande, and P. Topalova, "Female Leadership Raises Aspirations and Educational Attainment for Girls: A Policy Experiment in India," *Science*, vol. 335, no. 6068, pp. 582–586, 2012.

[62] R. Chattopadhyay and E. Duflo, "Women as policy makers: Evidence from a randomized policy experiment in India," *Econometrica*, vol. 72, no. 5, pp. 1409–1443, 2004.

[63] A. Clayton, "Women's political engagement under quota-mandated female representation: Evidence from a randomized policy experiment," *Comparative Political Studies*, vol. 48, no. 3, pp. 333–369, 2015.

[64] P. Zetterberg, "Do gender quotas foster women's political engagement? lessons from latin america," *Political Research Quarterly*, vol. 62, no. 4, pp. 715–730, 2009.

[65] T. J. Zagenczyk, R. Gibney, C. Kiewitz, and S. L. D. Restubog, "Mentors, supervisors and role models: do they reduce the effects of psychological contract breach?," *Human Resource Management Journal*, vol. 19, no. 3, pp. 237–259, 2009.

[66] T. Fried and A. MacCleave, "Influence of Role Models and Mentors on Female Graduate Students' Choice of Science as a Career," *Alberta Journal of Educational Research*, vol. 55, no. 4, pp. 482–496, 2009.

[67] B. Carrington and C. Skelton, "Re-thinking 'role models': equal opportunities in teacher recruitment in England and Wales," *Journal of Education Policy*, vol. 18, no. 3, pp. 253–265, 2003.

[68] E. P. Bettinger and B. T. Long, "Do Faculty Serve as Role Models? The Impact of Instructor Gender on Female Students," *The American Economic Review*, vol. 95, no. 2, pp. 152–157, 2005.

[69] C. L. Hoyt, "Inspirational or self-deflating. The role of self-efficacy in elite role model effectiveness," *Social Psychological and Personality Science*, vol. 4, no. 3, pp. 290–298, 2013.

[70] W. B. Mendes, J. Blascovich, B. Major, and M. Seery, "Challenge and threat responses during downward and upward social comparisons," *European Journal of Social Psychology*, vol. 31, no. 5, pp. 477–497, 2001.

[71] J. W. Osborne, "Testing stereotype threat: Does anxiety explain race and sex differences in achievement?," *Contemporary Educational Psychology*, vol. 26, no. 3, pp. 291–310, 2001.

[72] T. Schmader, "Gender Identification Moderates Stereotype Threat Effects on Women's Math Performance," *Journal of Experimental Social Psychology*, vol. 38, no. 2, pp. 194–201, 2002.

[73] J. Aronson, M. J. Lustina, C. Good, K. Keough, C. M. Steele, and J. Brown, "When white men can't do math: Necessary and sufficient factors in stereotype threat," *Journal of Experimental Social Psychology*, vol. 35, no. 1, pp. 29–46, 1999.

[74] J. K. Bosson, E. L. Haymovitz, and E. C. Pinel, "When saying and doing diverge: The effects of stereotype threat on self-reported versus non-verbal anxiety," *Journal of Experimental Social Psychology*, vol. 40, no. 2, pp. 247–255, 2004.

[75] K. Mogg and B. P. Bradley, "A cognitive-motivational analysis of anxiety," *Behaviour Research and Therapy*, vol. 36, no. 9, pp. 809–848, 1998.

[76] N. A. Valentino, V. L. Hutchings, A. J. Banks, and A. K. Davis, "Is a worried citizen a good citizen? emotions, political information seeking, and learning via the internet," *Political Psychology*, vol. 29, no. 2, pp. 247–273, 2008.

[77] J. Blascovich, S. J. Spencer, D. Quinn, and C. Steele, "African americans and high blood pressure: The role of stereotype threat," *Psychological Science*, vol. 12, no. 3, pp. 225–229, 2001.

[78] J. B. Hickam, W. H. Cargill, and A. Golden, "Cardiovascular reactions to emotional stimuli. effect on the cardiac output, arteriovenous oxygen difference, arterial pressure, and peripheral resistance," *Journal of Clinical Investigation*, vol. 27, no. 2, p. 290, 1948.

[79] J.-C. Croizet, G. Després, M.-E. Gauzins, P. Huguet, J.-P. Leyens, and A. Méot, "Stereotype threat undermines intellectual performance by triggering a disruptive mental load," *Personality and Social Psychology Bulletin*, vol. 30, no. 6, pp. 721–731, 2004.

[80] S. S. Townsend, B. Major, C. E. Gangi, and W. B. Mendes, "From "in the air" to "under the skin": Cortisol responses to social identity threat," *Personality and Social Psychology Bulletin*, vol. 37, no. 2, pp. 151–164, 2011.

[81] P. D. MacIntyre and R. C. Gardner, "Methods and results in the study of anxiety and language learning: A review of the literature," *Language Learning*, vol. 41, no. 1, pp. 85–117, 1991.

[82] V. Hodapp, P. G. Glanzmann, and L. Laux, "Theory and measurement of test anxiety as a situation-specific trait.," 1995.

[83] R. E. Smith, F. L. Smoll, and R. W. Schutz, "Measurement and correlates of sport-specific cognitive and somatic trait anxiety: The sport anxiety scale," *Anxiety Research*, vol. 2, no. 4, pp. 263–280, 1990.

[84] S. Kapitanoff and C. Pandey, "Stereotype threat, anxiety, instructor gender, and underperformance in women," *Active Learning in Higher Education*, vol. 18, no. 3, pp. 213–229, 2017.

[85] M. Cadinu, A. Maass, A. Rosabianca, and J. Kiesner, "Why do women underperform under stereotype threat? evidence for the role of negative thinking," *Psychological Science*, vol. 16, no. 7, pp. 572–578, 2005.

[86] M. Inzlicht and S. K. Kang, "Stereotype threat spillover: how coping with threats to social identity affects aggression, eating, decision making, and attention.," *Journal of Personality and Social Psychology*, vol. 99, no. 3, p. 467, 2010.

[87] A. Iyer, T. Schmader, and B. Lickel, "Why individuals protest the perceived transgressions of their country: The role of anger, shame, and guilt," *Personality and Social Psychology Bulletin*, vol. 33, no. 4, pp. 572–587, 2007.

[88] L. Huddy, L. Mason, and L. Aarøe, "Expressive partisanship: Campaign involvement, political emotion, and partisan identity," *American Political Science Review*, vol. 109, no. 1, pp. 1–17, 2015.

[89] C. A. Smith and L. D. Kirby, "Putting appraisal in context: Toward a relational model of appraisal and emotion," *Cognition and Emotion*, vol. 23, no. 7, pp. 1352–1372, 2009.

[90] R. J. Davidson, P. Ekman, C. D. Saron, J. A. Senulis, and W. V. Friesen, "Approach-withdrawal and cerebral asymmetry: Emotional expression and brain physiology: I.," *Journal of Personality and Social Psychology*, vol. 58, no. 2, p. 330, 1990.

[91] H. Kitschelt and D. M. Kselman, "Economic development, democratic experience, and political parties' linkage strategies," *Political Behavior*, vol. 46, no. 11, pp. 1453–1484, 2013.

[92] S.-H. Kim, D. A. Scheufele, and J. Shanahan, "Who cares about the issues? issue voting and the role of news media during the 2000 us presidential election," *Journal of Communication*, vol. 55, no. 1, pp. 103–121, 2005.

[93] J. G. Bullock, "Elite influence on public opinion in an informed electorate," *American Political Science Review*, vol. 105, no. 3, pp. 496–515, 2011.

[94] M. P. Fiorina, "The voting decision: instrumental and expressive aspects," *The Journal of Politics*, vol. 38, no. 2, pp. 390–413, 1976.

[95] J. A. Krosnick, "Government policy and citizen passion: A study of issue publics in contemporary america," *Political Behavior*, vol. 12, no. 1, pp. 59–92, 1990.

[96] S. Iyengar, K. S. Hahn, J. A. Krosnick, and J. Walker, "Selective exposure to campaign communication: The role of anticipated agreement and issue public membership," *The Journal of Politics*, vol. 70, no. 1, pp. 186–200, 2008.

[97] A. Bor, "Spontaneous categorization along competence in partner and leader evaluations," *Evolution and Human Behavior*, vol. 38, no. 4, pp. 468–473, 2017.

[98] M. C. Schneider and A. L. Bos, "Measuring stereotypes of female politicians," *Political Psychology*, vol. 35, no. 2, pp. 245–266, 2014.

[99] K. Dolan, "Gender stereotypes, candidate evaluations, and voting for women candidates: what really matters?," *Political Research Quarterly*, vol. 67, no. 1, pp. 96–107, 2014.

[100] K. Dolan and T. Lynch, "It takes a survey: Understanding gender stereotypes, abstract attitudes, and voting for women candidates," *American Politics Research*, vol. 42, no. 4, pp. 656–676, 2014.

[101] E. C. Cassese and M. R. Holman, "Party and gender stereotypes in campaign attacks," *Political Behavior*, pp. 1–23, 2017.

[102] N. J. Winter, "Masculine republicans and feminine democrats: Gender and americans' explicit and implicit images of the political parties," *Political Behavior*, vol. 32, no. 4, pp. 587–618, 2010.

[103] K. Sanbonmatsu, "Gender stereotypes and vote choice," *American Journal of Political Science*, pp. 20–34, 2002.

[104] L. Huddy and N. Terkildsen, "Gender stereotypes and the perception of male and female candidates," *American Journal of Political Science*, pp. 119–147, 1993.

[105] J. L. Lawless, "Women, war, and winning elections: Gender stereotyping in the post-september 11th era," *Political Research Quarterly*, vol. 57, no. 3, pp. 479–490, 2004.

[106] N. M. Bauer, "Rethinking stereotype reliance," *Politics and the Life Sciences*, vol. 32, no. 1, pp. 22–42, 2013.

[107] K. Dolan and K. Sanbonmatsu, "Gender stereotypes and attitudes toward gender balance in government," *American Politics Research*, vol. 37, no. 3, pp. 409–428, 2009.

[108] E. Perrault, "Why does board gender diversity matter and how do we get there? the role of shareholder activism in deinstitutionalizing old boys' networks," *Journal of Business Ethics*, vol. 128, no. 1, pp. 149–165, 2015.

[109] N. L. Erhardt, J. D. Werbel, and C. B. Shrader, "Board of director diversity and firm financial performance," *Corporate governance: An international review*, vol. 11, no. 2, pp. 102–111, 2003.

[110] S. Bear, N. Rahman, and C. Post, "The impact of board diversity and gender composition on corporate social responsibility and firm reputation," *Journal of Business Ethics*, vol. 97, no. 2, pp. 207–221, 2010.

[111] T. Nguyen, S. Locke, and K. Reddy, "Does boardroom gender diversity matter? evidence from a transitional economy," *International Review of Economics & Finance*, vol. 37, pp. 184–202, 2015.

[112] I. M. Young, *Inclusion and democracy*. Oxford University Press, 2002.

[113] N. Scherer and B. Curry, "Does descriptive race representation enhance institutional legitimacy? the case of the us courts," *The Journal of Politics*, vol. 72, no. 1, pp. 90–104, 2010.

[114] L. Huddy and N. Terkildsen, "Gender Stereotypes and the Perception of Male and Female Candidates," *American Journal of Political Science*, vol. 37, no. 1, pp. 119–147, 1993.

[115] K. Celis, "Substantive representation of women: the representation of women's interests and the impact of descriptive representation in the belgian parliament (1900–1979)," *Journal of Women, Politics & Policy*, vol. 28, no. 2, pp. 85–114, 2007.

[116] S. Childs and J. Lovenduski, "Political representation," in *The Oxford Handbook of Gender and Politics* (G. Waylen, ed.), Oxford University Press, 2013.

[117] K. A. Bratton, "Critical mass theory revisited: The behavior and success of token women in state legislatures," *Politics & Gender*, vol. 1, no. 1, pp. 97–125, 2005.

[118] WE, "Manifesto General Election 2017." Accessed at http://www.womensequality.org.uk/manifesto.

[119] K. Celis, S. Childs, J. Kantola, and M. L. Krook, "Constituting women's interests through representative claims," *Politics & Gender*, vol. 10, no. 2, pp. 149–174, 2014.

[120] R. Campbell, S. Childs, and J. Lovenduski, "Do women need women representatives?," *British Journal of Political Science*, vol. 40, no. 1, pp. 171–194, 2010.

[121] P. Cowley, "Why not ask the audience? understanding the public's representational priorities," *British Politics*, vol. 8, no. 2, pp. 138–163, 2013.

[122] H. F. Pitkin, *The Concept of Representation*. University of California Press, 1967.

[123] S. Dovi, *The Good Representative*. Malden, MA: John Wiley & Sons, 1 edition ed., 2007.

[124] E. A. Plant, P. G. Devine, W. T. Cox, C. Columb, S. L. Miller, J. Goplen, and B. M. Peruche, "The obama effect: Decreasing implicit prejudice and stereotyping," *Journal of Experimental Social Psychology*, vol. 45, no. 4, pp. 961–964, 2009.

[125] Parliament's Digital Service, "General Election 2015 Results." Accessed at http://www.data.parliament.uk/dataset/general-election-2015.

[126] D. Sanders and P. Whiteley, "British election study, 2010: Face-to-face survey." http://doi.org/10.5255/UKDA-SN-7529-1.

[127] D. Sanders and P. Whiteley, "British election study, 2010: Campaign internet data." http://doi.org/10.5255/UKDA-SN-7530-1.

[128] E. Fieldhouse, J. Green, G. Evans, H. Schmitt, C. van der Eijk, J. Mellon, and C. Prosser, "British Election Study Internet Panel Waves 1-6," 2017. http://doi.org/10.5255/UKDA-SN-8202-1.

[129] J. M. Mendez and R. Herrick, *Descriptive Elections: Empowering the American Electorate*. Lexington Books, 2016.

[130] S. Verba, N. Burns, and K. L. Schlozman, "Knowing and Caring about Politics: Gender and Political Engagement," *The Journal of Politics*, vol. 59, no. 4, pp. 1051–1072, 1997.

[131] S. Childs and M. L. Krook, "Critical mass theory and women's political representation," *Political Studies*, vol. 56, no. 3, pp. 725–736, 2008.

[132] D. T. Studlar and I. McAllister, "Does a critical mass exist? a comparative analysis of women's legislative representation since 1950," *European Journal of Political Research*, vol. 41, no. 2, pp. 233–253, 2002.

[133] SMI, "Begaze manual," 2011. Accessed at
http://twiki.cis.rit.edu/twiki/pub/MVRL/SmiTracker/begaze2.pdf.

[134] D. Stevens, S. Banducci, G. Katz Wisel, T. Coan, I. Cioroianu, E. Kolpinskaya,
and J. Ault, "Media in context," 2015. Accessed at
https://mediaeffectsresearch.wordpress.com.

[135] E. H. Koster, G. Crombez, B. Verschuere, and J. De Houwer, "Selective attention
to threat in the dot probe paradigm: Differentiating vigilance and difficulty to
disengage," *Behaviour Research and Therapy*, vol. 42, no. 10, pp. 1183–1192,
2004.

[136] S. R. Staugaard, "Reliability of two versions of the dot-probe task using
photographic faces," *Psychology Science Quarterly*, vol. 51, no. 3, pp. 339–350,
2009.

[137] J. S. Lerner, R. M. Gonzalez, D. A. Small, and B. Fischhoff, "Effects of fear and
anger on perceived risks of terrorism: A National field experiment,"
Psychological Science, vol. 14, no. 2, pp. 144–150, 2003.

[138] J. S. Lerner, D. A. Small, and G. Loewenstein, "Heart strings and purse strings
carryover effects of emotions on economic decisions," *Psychological Science*,
vol. 15, no. 5, pp. 337–341, 2004.

[139] T. Coan, J. Merolla, and E. J. Zechmeister, "Emotional responses to human
security threats: Evidence from a national experiment." Accessed at
scholar.harvard.edu/files/tcoan/files/emotions.pdf. Manuscript, accessed
23/02/2016.

[140] J. W. Pennebaker, R. J. Booth, and M. E. Francis, "LIWC: Linguistic inquiry and
word count," 2007. Accessed at
http://www.liwc.net/LIWC2007LanguageManual.pdf.

[141] Y. R. Tausczik and J. W. Pennebaker, "The psychological meaning of words:
Liwc and computerized text analysis methods," *Journal of Language and Social
Psychology*, vol. 29, no. 1, pp. 24–54, 2010.

[142] E. Peer, L. Brandimarte, S. Samat, and A. Acquisti, "Beyond the turk: Alternative
platforms for crowdsourcing behavioral research," *Journal of Experimental Social
Psychology*, vol. 70, pp. 153–163, 2017.

[143] R. Dassonneville and I. McAllister, "Gender, political knowledge, and descriptive
representation: The impact of long-term socialization," *American Journal of
Political Science*, 2018. Early View.

[144] J. E. Filer and L. W. Kenny, "Voter turnout and the benefits of voting," *Public
Choice*, vol. 35, no. 5, pp. 575–585, 1980.

[145] J. Lovenduski, "Gender politics: A breakthrough for women?," *Parliamentary
Affairs*, vol. 50, no. 4, pp. 708–719, 1997.

[146] R. R. Lau, "Information search during an election campaign: Introducing a
processing-tracing methodology for political scientists," in *Political Judgment:
Structure and Process* (M. Lodge and K. M. McGraw, eds.), ch. 8, pp. 179–206,
University of Michigan Press, 1995.

[147] L. Carraro, L. Castelli, and C. Macchiella, "The automatic conservative: Ideology-based attentional asymmetries in the processing of valenced information," *PLoS One*, vol. 6, no. 11, 2011.

[148] Z. Wang, A. C. Morey, and J. Srivastava, "Motivated selective attention during political ad processing: The dynamic interplay between emotional ad content and candidate evaluation," *Communication Research*, vol. 41, no. 1, pp. 119–156, 2014.

[149] C. MacLeod, A. Mathews, and P. Tata, "Attentional bias in emotional disorders.," *Journal of Abnormal Psychology*, vol. 95, no. 1, p. 15, 1986.

[150] M. A. Just and P. A. Carpenter, "Eye fixations and cognitive processes," *Cognitive Psychology*, vol. 8, no. 4, pp. 441–480, 1976.

[151] M. Longair, S. Roe, Z. Zappia, *et al.*, "Democracy Club Candidate Data," 2018. Accessed at https://candidates.democracyclub.org.uk/help/api.

[152] M. Prior, "You've either got it or you don't? The stability of political interest over the life cycle," *The Journal of Politics*, vol. 72, no. 3, pp. 747–766, 2010.

[153] M. X. D. Carpini and S. Keeter, *What Americans know about politics and why it matters*. Yale University Press, 1996.

[154] V. Larcinese, "Does political knowledge increase turnout? evidence from the 1997 british general election," *Public Choice*, vol. 131, no. 3-4, pp. 387–411, 2007.

[155] Electoral Commission and The Hansard Society, "An audit of political engagement," 2005. Accessed at https://www.hansardsociety.org.uk/research/audit-of-political-engagement.

[156] M. J. Shih, T. L. Pittinsky, and G. C. Ho, "Stereotype boost: Positive outcomes from the activation of positive stereotypes.," in *Stereotype threat: Theory, process, and application* (M. Inzlicht and T. Schmader, eds.), pp. 141–156, Oxford University Press, 2012.

[157] J. L. Smith and C. S. Johnson, "A stereotype boost or choking under pressure? positive gender stereotypes and men who are low in domain identification," *Basic and Applied Social Psychology*, vol. 28, no. 1, pp. 51–63, 2006.

[158] T. Clark, "Nervous nellies. Girls don't start out more anxious than boys, but they usually end up that way.," *Slate*, 2011. Accessed at 22 March 2018, at http://www.slate.com/human-interest/2018/03/our-one-fight-the-scorekeeper-vs-the-pitcher-inner.html.

[159] S. A. Banducci, T. Donovan, and J. A. Karp, "Minority Representation, Empowerment, and Participation," *The Journal of Politics*, vol. 66, no. 2, pp. 534–556, 2004.

[160] M. Johns, T. Schmader, and A. Martens, "Knowing is half the battle: Teaching stereotype threat as a means of improving women's math performance," *Psychological Science*, vol. 16, no. 3, pp. 175–179, 2005.

[161] L. T. O'Brien and C. S. Crandall, "Stereotype threat and arousal: Effects on women's math performance," *Personality and Social Psychology Bulletin*, vol. 29, no. 6, pp. 782–789, 2003.

[162] R. M. Baron and D. A. Kenny, "The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations." *Journal of Personality and Social Psychology*, vol. 51, no. 6, p. 1173, 1986.

[163] D. D. Rucker, K. J. Preacher, Z. L. Tormala, and R. E. Petty, "Mediation analysis in social psychology: Current practices and new recommendations," *Social and Personality Psychology Compass*, vol. 5, no. 6, pp. 359–371, 2011.

[164] D. P. MacKinnon, A. J. Fairchild, and M. S. Fritz, "Mediation analysis," *Annual Review of Psychology*, vol. 58, pp. 593–614, 2007.

[165] J. G. Bullock and S. E. Ha, "35. Mediation Analysis Is Harder than It Looks," 2011.

[166] C. D. Spielberger, *State-Trait anxiety inventory*. Wiley Online Library, 2010.

[167] R. J. Mokken, *A theory and procedure of scale analysis: With applications in political research*, vol. 1. Walter de Gruyter, 1971.

[168] L. A. Van der Ark *et al.*, "Mokken scale analysis in r," *Journal of Statistical Software*, vol. 20, no. 11, pp. 1–19, 2007.

[169] L. A. Van der Ark, M. A. Croon, and K. Sijtsma, "Mokken scale analysis for dichotomous items using marginal models," *Psychometrika*, vol. 73, no. 2, p. 183, 2008.

[170] S. E. Embretson and S. P. Reise, *Item response theory*. Psychology Press, 2013.

[171] D. Rizopoulos, "ltm: An r package for latent variable modeling and item response theory analyses," *Journal of Statistical Software*, vol. 17, no. 5, pp. 1–25, 2006.

[172] K. G. Lynch, M. Cary, R. Gallop, and T. R. Ten Have, "Causal mediation analyses for randomized trials," *Health Services and Outcomes Research Methodology*, vol. 8, no. 2, pp. 57–76, 2008.

[173] C. E. Frangakis and D. B. Rubin, "Principal stratification in causal inference," *Biometrics*, vol. 58, no. 1, pp. 21–29, 2002.

[174] B. Albertson and S. K. Gadarian, "Did that scare you? tips on creating emotion in experimental subjects," *Political Analysis*, vol. 24, no. 4, pp. 485–491, 2016.

[175] B. Pang, L. Lee, *et al.*, "Opinion mining and sentiment analysis," *Foundations and Trends® in Information Retrieval*, vol. 2, no. 1–2, pp. 1–135, 2008.

[176] J. Pennebaker, C. Chung, M. Ireland, A. Gonzales, and R. Booth, "The development and psychometric properties of LIWC 2007.[software manual]," 2007.

[177] R. L. Nabi, "Exploring the framing effects of emotion: do discrete emotions differentially influence information accessibility, information seeking, and policy preference?," *Communication Research*, vol. 30, no. 2, pp. 224–247, 2003.

[178] M. Van Zomeren, R. Spears, A. H. Fischer, C. W. Leach, *et al.*, "Put your money where your mouth is! explaining collective action tendencies through group-based anger and group efficacy," *Journal of Personality and Social Psychology*, vol. 87, pp. 649–664, 2004.

[179] P. Tough, "Obama Effect, The," *The New York Times*, 2009. Accessed at 18 October 2017, at https://nyti.ms/2zjg5Hh.

[180] B. Rigby, "Theresa May is not so different from David Cameron after all," *Sky Views*. Accessed at http://news.sky.com/story/sky-views-theresa-may-can-make-the-tories-appealing-to-women-10900204.

[181] P. Gurin, "Women's gender consciousness," *Public Opinion Quarterly*, vol. 49, no. 2, pp. 143–163, 1985.

[182] L. Huddy and N. Terkildsen, "The consequences of gender stereotypes for women candidates at different levels and types of office," *Political Research Quarterly*, vol. 46, no. 3, pp. 503–525, 1993.

[183] K. Dolan, "The impact of gender stereotyped evaluations on support for women candidates," *Political Behavior*, vol. 32, no. 1, pp. 69–88, 2010.

[184] A. Borchorst and B. Siim, "Woman-friendly policies and state feminism: Theorizing scandinavian gender equality," *Feminist Theory*, vol. 9, no. 2, pp. 207–224, 2008.

[185] A. Gelman, A. Jakulin, M. G. Pittau, and Y.-S. Su, "A weakly informative default prior distribution for logistic and other regression models," *The Annals of Applied Statistics*, pp. 1360–1383, 2008.

[186] J. K. Kampen, S. V. De Walle, and G. Bouckaert, "Assessing the relation between satisfaction with public service delivery and trust in government. the impact of the predisposition of citizens toward government on evalutations of its performance," *Public Performance & Management Review*, vol. 29, no. 4, pp. 387–404, 2006.

[187] J. D. Hadfield *et al.*, "Mcmc methods for multi-response generalized linear mixed models: the mcmcglmm r package," *Journal of Statistical Software*, vol. 33, no. 2, pp. 1–22, 2010.

[188] L. A. Schwindt-Bayer and C. Reyes-Housholder, "Citizen responses to female executives: is it sex, novelty or both?," *Politics, Groups, and Identities*, vol. 5, no. 3, pp. 373–398, 2017.

[189] P. C. Flore and J. M. Wicherts, "Does stereotype threat influence performance of girls in stereotyped domains? a meta-analysis," *Journal of school psychology*, vol. 53, no. 1, pp. 25–44, 2015.

[190] F. G. F. Foos and F. Gilardi, "Bridging the gender gap in political ambition: Experimental evidence," 2016.

[191] B. L. Fraga, "Redistricting and the causal impact of race on voter turnout," *The Journal of Politics*, vol. 78, no. 1, pp. 19–34, 2016.

[192] J. A. Henderson, J. S. Sekhon, and R. Titiunik, "Cause or effect? turnout in hispanic majority-minority districts," *Political Analysis*, vol. 24, no. 3, pp. 404–412, 2016.