

What underlies the opposition to trans-inclusive policies? The role of concerns about male violence versus attitudes towards trans people

Thekla Morgenroth <sup>1</sup>

Jordan R. Axt <sup>2,3</sup>

Erin C. Westgate <sup>4</sup>

<sup>1</sup> Purdue University

<sup>2</sup> McGill University

<sup>3</sup> Project Implicit

<sup>4</sup> University of Florida

**Author Contact Information:**

Thekla Morgenroth  
Department of Psychological Sciences, Room 2160  
703 Third Street  
West Lafayette, Indiana 47907  
tmorgenr@purdue.edu

## Abstract

Transgender women's access to women-only spaces is controversial. Arguments against trans-inclusive policies often focus on cisgender women's safety from male violence, despite little evidence to suggest that such policies put cisgender women at risk. Across seven studies using US and UK participants ( $N=3864$ ), we investigate whether concerns about male violence versus attitudes toward trans people are a better predictor of support for trans-inclusive policies, and whether these factors align with the reasons given by opponents and supporters regarding their policy views. We find that opponents of these policies do not accurately report their reasons for opposition: specifically, while opponents claim that concerns about male violence are the primary reason driving their opposition, attitudes towards transgender people more strongly predicted policy views. These results highlight the limitations of focusing on overt discourse and emphasize the importance of investigating psychological mechanisms underlying policy support.

*Keywords:* Policy, transphobia, transgender, implicit attitudes, gender-based violence

What underlies the opposition to trans-inclusive policies? The role of concerns about male violence versus attitudes towards trans people

*“I want trans women to be safe. At the same time, I do not want to make natal girls and women less safe. When you throw open the doors of bathrooms and changing rooms to any man who believes or feels he’s a woman, [...] then you open the door to any and all men who wish to come inside.”* (Rowling, 2020)

The quote above, by *Harry Potter* author J. K. Rowling, illustrates a common argument made in the context of trans-inclusive policies, particularly policies regarding transgender women’s (i.e. women whose gender identity does not match their sex assigned at birth) access to women-only spaces: namely, that trans-inclusive policies place cisgender women (i.e., women whose gender identity matches their sex assigned at birth) at danger from male violence. At first glance, such arguments may seem convincing. Women and girls are the primary targets of sexual assault, while men are its primary perpetrators (NISVS, 2010; Office of National Statistics, 2017), and it is reasonable to demand policies that increase women and girls’ safety. However, these safety-based arguments hold little empirical weight. For instance, the vast majority of sexual violence is committed in private spaces by perpetrators previously known to the victim, not by strangers in public spaces (NISVS, 2010). Additionally, evidence fails to show that trans-inclusive policies reduce safety or privacy; one comparison of localities in Massachusetts that passed (or did not pass) such ordinances showed no change in reported privacy or safety violations in public restrooms, locker rooms, or dressing rooms (Hasenbush et al., 2019).

It is of course possible that opponents of trans-inclusive policies are not aware of these findings and that their desire to protect cisgender women from male violence indeed drives their opposition. We argue, however, that opposition to such policies is driven by negative attitudes towards transgender people and that concerns for cisgender women’s safety are used as a seemingly valid rationale to oppose trans-inclusive policies while allowing one

to appear non-prejudiced. Importantly, this reliance on safety concerns as justification for opposition may be used not just to deceive others but also to deceive oneself.

A long history of psychological research shows that people often lack conscious access to the reasons underlying their behaviors and beliefs (e.g., Nisbett & Wilson, 1977; Wilson, 2004), with the result that people may come to endorse inaccurate – and often motivated - explanations. For instance, people endure unpleasant experiences longer when told that endurance is associated with desirable traits such as good skin, but deny that this behavior is due to greater effort (Fernbach et al., 2014), and cheat on practice problems when told an upcoming exam is diagnostic of future success but deny having cheated, thus convincing themselves that their ability is higher than it actually is (Paulhus & Buckles, 2012). Such self-deception may also occur when people are motivated to appear nonprejudiced. For instance, people who profess egalitarian beliefs are more likely to question the validity of measures of implicit associations if their own performance indicates bias (Howell et al, 2017).

Similar processes might be at play in beliefs concerning the treatment of transgender people and prior work supports our argument that attitudes towards transgender people (rather than professed concern about the threat of male violence) is more strongly associated with opposition to trans-inclusive policies. First, explicit and implicit transgender attitudes predict policy support, including support for trans-inclusive policies (Axt et al., 2020), and such anti-trans attitudes appear to be widespread and robust (Axt et al., 2020). At the same time, societal values largely condemn prejudice (Crandall et al., 2002) and individuals are motivated to control and suppress their own prejudice to preserve a positive self-image (Crandall & Eshleman, 2003).

When faced with decisions regarding trans-inclusive policies, people with negative attitudes towards transgender people thus face a dilemma. On the one hand, they may feel

inclined to oppose such policies based on their prejudice, but on the other hand, they may feel pressure to not appear prejudiced. To solve this dilemma, our minds have developed several strategies to justify prejudiced actions, including *covering* (see Crandall & Eshleman, 2003). Covering refers to the process of obscuring underlying prejudice by providing an alternative explanation for prejudiced actions. Situational ambiguity (i.e., a range of different potential explanations for an action, some of them neutral or benign) makes this strategy particularly likely (Crandall & Eshleman, 2003). For example, right-wing politicians often distinguish between Muslims (as individuals they have no negative attitudes towards) and Islam (as a harmful ideology that threatens Western values), thus justifying their anti-Muslim policies while covering their prejudice (Verkuyten, 2013). The logic of this argument is similar to the one made by J. K. Rowling, cited at the beginning of this paper, where she argues that her issue lies not with trans women (i.e., the target of trans-inclusive policies), but with a harmful set of beliefs that threaten women's safety.

We argue that a similar mechanism is at play here. Because concern for cisgender women's safety appears benign (or even noble), it is reported as a justification for opposition to trans-inclusive policies to cover up prejudice that is more strongly related to policy opposition. Because people high in explicit prejudice may be less motivated to obscure their true motivation, we also explored the possibility that implicit prejudice affects the extent to which individuals draw on male violence as a reason for their opposition.

### **The Current Project**

We report five correlational and two experimental studies investigating whether concerns about male violence (versus attitudes toward trans people) are a better predictor of support for trans-inclusive policies, and whether these factors align with the reasons cited by policy opponents and supporters.

Across five pre-registered correlational studies, we asked people to report their support for trans-inclusive policies as well as their reasons for supporting (or opposing) them. We then compared self-reported motives to actual observed associations between policy support and 1) attitudes towards transgender people and 2) beliefs about male violence, respectively. In two pre-registered experiments, we aimed to manipulate male violence beliefs and attitudes towards transgender people directly.

In Studies 1-5, we predicted that opponents of trans-inclusive policies would portray their reasons for support/opposition less accurately than policy supporters (Hypothesis 1). More specifically, we hypothesized that opponents would cite their belief that men are violent as more relevant to their policy positions, while supporters would cite their attitudes toward trans people (Hypothesis 1a). However, we predicted that - for both opponents and supporters - attitudes toward trans people (implicit and explicit) would reliably predict trans-inclusive policy support/opposition, but the belief that men are a threat to women would not. As a result, transgender attitudes (both explicit and implicit) should predict support more strongly than the self-reported association of men with violence (Hypothesis 1b).<sup>1</sup>

These predictions rest on the assumption that a stronger correlation with policy positions indicates greater relevance. This model of inference is common in the social sciences; for example, Abramowitz (1995)'s assertion that abortion is the dominant issue in understanding voting behavior stems from evidence that abortion positions were a stronger predictor of presidential vote than any other policy belief, such as views on the death penalty, or affirmative action. Similarly, we compare the predictive value of transgender attitudes versus beliefs about male violence for understanding people's positions on trans-inclusive policies. Such approaches are often necessary in domains where people may be unwilling or

---

<sup>1</sup> We pre-registered different hypotheses for Study 1 and an additional hypothesis for Studies 2-5. These hypotheses and respective analyses can be found in the online supplement. Please also note that the wording of our hypotheses changed slightly across studies (see pre-registrations).

unable to accurately self-report what factors are responsible for their behavior (Crandall & Eshleman, 2003).

Because the methods of Studies 1-4 are fairly consistent, we present our studies meta-analytically to increase precision and robustness (Goh et al., 2016; for examples, see Kreps et al., 2017; Kirby et al., 2020). The analyses in Study 1 were largely exploratory, so Study 2 sought to replicate the findings in a larger US sample. Studies 3-4 were replications with samples from the UK using a more refined measure of policy support; Study 4 fixed a programming error present in Study 3.

Below, we first report the methods used in Studies 1-4, followed by the meta-analytic results for these studies. We then present Studies 5-7 as separate studies. Study 5 employed different operationalizations of male violence and transgender prejudice to ensure that findings were not specific to the measures used in Studies 1-4. Finally, Studies 6-7 used experimental designs to enable us to make stronger causal claims.

Full materials and data for all studies are available at [https://osf.io/7qjgt/?view\\_only=f8bb2c674b6c458997aef516c5f7a7b7](https://osf.io/7qjgt/?view_only=f8bb2c674b6c458997aef516c5f7a7b7). We report all measures and exclusions of our studies. Additional methodological details can be found in the online supplement.

### **Studies 1-4: Meta-Analysis of Reasons**

We conducted four studies asking opponents and supporters of trans-inclusive policies to report their reasons for their policy stances, and compared the stated reasons to their actual empirical predictors. Studies 1 and 2 included additional measures, which are reported in the online supplement.

## Method

### *Participants*

We recruited participants from the US and the UK as trans-inclusive policies are widely debated in both countries. Sample sizes for individual studies were based on financial constraints and, for Studies 2-4, on effect sizes from previous studies. Demographic information for all studies can be found in Table 1.

**Table 1**  
*Demographic Information for all Studies*

	Study 1	Study 2	Study 3	Study 4	Study 5	Study 6	Study 7
Nationality	US	US	UK	UK	US	US	US
Gender							
Women	49.25%	62.09%	55.23%	58.67%	67.26%	53.72%	46.57%
Men	48.26%	37.25%	44.77%	41.33%	32.74%	44.99%	51.51%
Race							
Asian	14.93%	4.58%	-	-	6.61%	8.92%	6.59%
Black	5.97%	10.46%	-	-	14.64%	6.02%	8.65%
Middle Eastern	-	-	-	-	1.07%	0.72%	0.96%
Native American	-	-	-	-	1.07%	1.72%	2.61%
Pacific Islander	-	-	-	-	0.18%	0%	0%
White	67.66%	72.11%	-	-	57.52%	81.81%	80.22%
Hispanic	-	-	-	-	12.32%	5.44%	7.01%
LGBTQ+	-	14.38%	-	-	12.92%	-	-
Age ( <i>SD</i> )	33.02 (13.21)	40.72 (15.28)	36.48 (14.43)	38.77 (15.29)	33.16 (14.50)	41.47 (14.35)	41.33 (14.36)
Political ideology ( <i>SD</i> )	-	-	3.59 (1.67)	3.58 (1.54)	0.55 (1.73)	-	-

*Note.* Blank cells indicate that the information was not collected in that study. Percentages may not add up to 100% because of missing information or because participants could select multiple categories. Political ideology was measured on a scale from 1 (very progressive) to 7 (very conservative) in Studies 3-4, and on a scale from -3 to +3 in Study 5.

**Study 1.** We recruited a preregistered sample of 230 participants via Prolific (<https://aspredicted.org/blind.php?x=5qt3uw>). After excluding 29 participants who withdrew their consent, were transgender, did not indicate they were US American, were under 18, or were missing IAT scores, our final sample size was 201, yielding 80% power for H1a to detect an effect size of  $f=.14$ .

**Study 2.** We recruited a pre-registered ([https://osf.io/6tw3a/?view\\_only=0cf5f0c46cd7479f9d2907f47ee3a04f](https://osf.io/6tw3a/?view_only=0cf5f0c46cd7479f9d2907f47ee3a04f)) sample of 450 Project Implicit volunteers, with 80% power to detect an effect size of  $f=.09$  (based on Study



1, and accounting for exclusions). Because we excluded fewer participants than anticipated (2 because they identified as transgender and 12 due to fast responding on the IAT), our final sample consisted of 459 cisgender US Americans.

**Study 3.** We recruited a preregistered sample of 740 British participants via Prolific (<https://aspredicted.org/blind.php?x=3u6wj2>), excluding participants who identified as non-binary or transgender, who were not UK citizens/residents, or who responded too quickly to 10% or more of the IAT items, yielding a final sample size of 717, with 80% power to detect an effect size of  $f=.06$ . We sought to sample equally across the political spectrum by recruiting separately (1/3 each) from participants who indicated that their political affiliation was “left,” “centre,” or “right.”

**Study 4.** We recruited a preregistered sample of 500 British participants via Prolific (<https://aspredicted.org/blind.php?x=pe44bs>), excluding participants who identified as non-binary or transgender or who responded too quickly to 10% or more of the IAT items, yielding a final sample of 496, with 80% power to detect an effect size of  $f=.08$ . We sought to sample equal numbers of participants across the political spectrum, using the same strategy as in Study 3.

### ***Procedure***

Our studies were advertised as being about predictors of policy support. Across all studies, participants completed a transgender attitudes Implicit Association Test (IAT; Greenwald et al., 1998), explicit measures of transgender attitudes and association of men with violence (order of explicit and implicit measures randomized), and a measure of support for trans-inclusive policies (either before or after all other measures). In Studies 1-2, we used the policy support measure to categorize participants as supporters (i.e., scoring above the midpoint) or opponents (i.e., scoring below the midpoint). Participants who scored at the midpoint were omitted from the analyses reported here (see online supplement for additional

details). In Studies 3 and 4, participants self-categorized as supporters or opponents of trans-inclusive policies prior to the policy support measure.

### **Measures**

**Transgender IAT.** We measured implicit attitudes towards transgender people using a validated seven-block image IAT (see Axt et al., 2020). In the IAT, participants sorted good (e.g., joyful, pleasure) and bad words (e.g., poison, rotten) into “good” and “bad” categories, and images of eight celebrities into transgender and cisgender categories. The celebrities were four transgender women and men (Chaz Bono, Chris Mosier, Laverne Cox, and Caitlyn Jenner) as well as four cisgender women and men (Jon Favreau, Tyler Clary, Meagan Goode, and Meryl Streep). Transgender and cisgender targets were matched on gender identity, race, age, and popularity. To ensure that participants were familiar with the celebrities, participants first read a brief statement about each celebrity and completed a training block sorting the images into transgender and cisgender categories. We used the *D* algorithm (Greenwald et al., 2003) to score responses, such that higher values indicate more positive attitudes towards cisgender versus transgender people.

**Explicit attitudes towards transgender people.** We used five items to measure explicit transgender attitudes (from Axt et al., 2020;  $\alpha$  ranging from .82 to .91). Participants reported their preference on a scale from 1 (“I strongly prefer transgender to cisgender people”) to 7 (“I strongly prefer cisgender to transgender people”), indicated how warm they feel towards transgender and cisgender people from 1 (very cold) to 7 (very warm), and how positive they felt towards transgender and cisgender people from 1 (very negative) to 7 (very positive). We then calculated a relative 7-point measure (with the exception of Study 2; see online supplement). Higher numbers indicate a stronger preference for cisgender people over transgender people.

**Gender-violence association.** Participants responded to five items about their relative associations between gender and violence, again forming a relative measure. First, participants indicated the extent to which they associated women and men with violence (versus peace) on a scale from 1 (“I strongly associate women with peace and men with violence”) to 7 (“I strongly associate men with peace and women with violence”; reverse-coded). Next, participants indicated their belief about which gender is more likely to commit a violent crime and resolve conflict peacefully and calmly on a scale from 1 (women, much more) to 7 (men, much more). Finally, participants reported the extent to which they associated the concepts of peaceful and violent with women versus men on a scale from 1 (women, strongly) to 7 (men, strongly). These items were highly face-valid, the format mirrored the explicit attitude measure (from Axt et al., 2020), and the resulting scale was reliable ( $\alpha$  ranging from .79 to .86). We calculated a relative measure on a scale from 1-7 (with the exception of Study 2), where higher numbers indicate a stronger association of men with violence and women with peace.

**Support for trans-inclusive policies.** Participants indicated their agreement with statements about four trans-inclusive policies: “Women’s shelters should also be accessible to transgender women,” “Transgender people should be able to use the bathroom of the gender they most closely identify with,” “If transgender people go to prison, they should go to the prison that is aligned with their sex assigned at birth,” (reverse-scored) and “Sex-segregated sports teams should allow transgender people to join the teams they feel most comfortable in,”  $\alpha$  ranging from .81 to .92) on a scale from “strongly disagree” to “strongly agree.” The scale was a seven-point scale in Study 1, a six-point scale (to force participants to choose between support and opposition) in Study 2, and an 11-point-scale (to generate greater variability within opponents and supporters) in Studies 3 and 4.

Items changed slightly across studies. For Study 2, we explicitly defined “transgender people” as based on gender identity regardless of whether or not individuals had undergone medical transition such as hormone replacement therapy or gender confirmation surgery. For Studies 3-4 we replaced the item “Sex-segregated sports teams should allow transgender people to join the teams they feel most comfortable in, regardless of whether they have received hormone replacement therapy or gender confirmation surgery” with the item “Transgender women should have access to women-only support groups for victims of sexual or domestic abuse” because open-ended participant responses suggested that factors not assessed here (such as fairness perceptions, rather than safety concerns) may play a role in the opposition to the inclusion of trans people in sex-segregated sports teams. These items were developed for the purpose of this project, are high in face validity, reflect actual policies, and formed reliable scales across our studies.

**Reasons for support and opposition.** Participants then indicated their reasons for support and opposition. In Study 1, they reported the extent to which 1) their feelings towards transgender people and 2) their belief that men are (or are not) a threat to women affected their agreement with the policy statements on a scale from 1 (not at all) to 7 (very strongly). In Study 2-4 we asked about reasons for support and opposition separately. More specifically, we asked them to what extent 1) their feelings towards trans people and 2) their thoughts around violence committed by men affected their support and opposition, respectively, on 7-point scales. For example, the words “My thoughts around violence committed by men...” were followed by a bipolar scale from “... had nothing to do with my opposition or support for the policies” to “led me to oppose the policies.” We created mean scores of the two trans attitudes ( $r$ s ranging from .29 to .40) and the two male violence items ( $r$ s ranging from .39 to .62) to calculate overall scores of the importance of the reasons (comparable to those in Study 1) to test whether opponents of trans-inclusive policies would portray their reasons less

accurately than policy supporters, but examined reasons for support and opposition separately when testing whether this was moderated by implicit trans attitudes.

## **Results**

### *Meta-analysis*

We predicted that opponents of trans-inclusive policies would portray their reasons for support/opposition less accurately than policy supporters. Descriptive statistics and zero-order correlations for supporters versus opponents respectively, are reported in Tables 2 and 3. Trans attitudes (explicit and implicit) were not related to male violence beliefs, making it unlikely that trans attitudes are informed by the belief that men are violent, or vice versa.

Table 2

*Descriptive Statistics for Opponents and Supporters of Trans-inclusive Policies (Studies 1-4)*

	Study 1				Study 2				Study 3				Study 4			
	Opponents		Supporters		Opponents		Supporters		Opponents		Supporters		Opponents		Supporters	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Trans attitudes IAT d score	0.40	0.43	0.24	0.46	0.29	0.36	0.19	0.42	0.51	0.42	0.24	0.45	0.51	0.41	0.29	0.44
Explicit trans attitudes	5.37	1.04	4.14	0.65	0.57	1.16	-0.18	0.66	4.66	0.91	4.04	0.49	4.61	0.85	4.03	0.44
Gender-violence association	5.11	0.95	5.32	0.80	-0.12	0.83	0.04	0.82	5.15	0.75	5.14	0.74	5.24	0.74	5.18	0.68
Support for trans-inclusive policies	2.39	0.94	5.47	0.88	2.40	0.72	5.09	0.73	3.94	2.25	8.62	2.08	3.90	2.18	8.31	2.24
Reason: Trans attitudes	3.84	2.06	4.30	1.61	2.49	1.53	3.84	1.74	2.89	1.59	3.45	1.52	2.85	1.51	3.34	1.46
Reason: Male violence	3.85	1.71	3.62	1.67	3.20	1.89	3.19	1.85	3.19	1.68	3.04	1.58	3.12	1.77	3.05	1.60

*Note.* Explicit trans attitudes and gender-violence association are displayed in standardized form for Study 2, but unstandardized (scale from 1-7) for all other studies. Support for trans-inclusive policies was measured on a scale from 1-7 in Study 1, on a scale from 1-6 in Study 2, and on a scale from 1-11 in Studies 3-4.

Table 3  
*Bivariate Correlations between Variables (Studies 1-4)*

	1	2	3	4	5	6
<i>Study 1</i>						
1. Trans attitudes IAT <i>D</i> score	-	.25**	.16	-.19*	-.01	.06
2. Explicit trans attitudes	.06	-	.00	-.26**	-.13	.22*
3. Explicit gender-violence association	.02	.15	-	.18	.13	.12
4. Support for trans-inclusive policies	-.24*	-.40**	-.13	-	.36***	-.25**
5. Reasons: Trans attitudes	-.11	.43***	.14	.04	-	.12
6. Reasons: Male violence	-.13	.00	.20	.03	-.03	-
<i>Study 2</i>						
1. Trans attitudes IAT <i>D</i> score	-	.163**	.04	-.13*	-.11*	.02
2. Explicit trans attitudes	.04	-	-.02	-.23***	-.17**	.01
3. Explicit gender-violence association	-.04	.14	-	.10	.16**	.2-***
4. Support for trans-inclusive policies	.15	-.16	-.04	-	.32***	-.03
5. Reasons: Trans attitudes	-.02	.13	.10	.01	-	.35***
6. Reasons: Male violence	.00	.02	.24*	.14	.20*	-
<i>Study 3</i>						
1. Trans attitudes IAT <i>D</i> score	-	.34***	-.04	-.17**	-.08	.06
2. Explicit trans attitudes	.29***	-	-.04	-.27***	-.02	.03
3. Explicit gender-violence association	.03	.05	-	.12*	.11*	.18**
4. Support for trans-inclusive policies	-.26***	-.39***	-.09	-	.05	-.20***
5. Reasons: Trans attitudes	.14*	.42***	.08	-.03	-	.30***
6. Reasons: Male violence	-.02	-.05	.15**	.07	.26***	-
<i>Study 4</i>						
1. Trans attitudes IAT <i>D</i> score	-	.26***	.01	-.19**	-.10	.03
2. Explicit trans attitudes	.23***	-	-.06	-.29***	-.08	-.04
3. Explicit gender-violence association	.02	.09	-	.18**	.15*	.18**
4. Support for trans-inclusive policies	-.05	-.30***	-.25***	-	.14*	-.17**
5. Reasons: Trans attitudes	.1	.28***	-.02	-.04	-	.23***
6. Reasons: Male violence	.01	-.13	.23***	-.01	.01	-

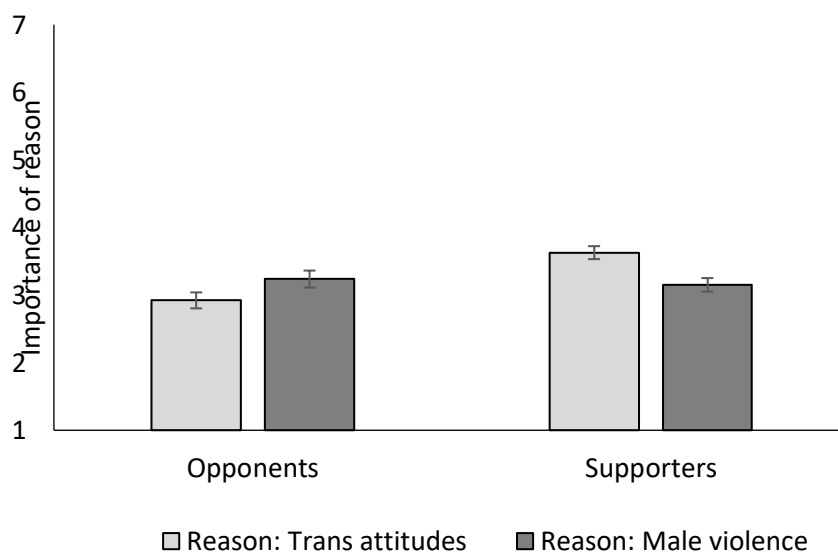
*Note.* \*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .001$ . Correlations above the diagonals refer to correlations for supporters of trans-inclusive policies and values below the diagonals refer to correlations for opponents of trans-inclusive policies.

**Professed reasons for policy support (Hypothesis H1a).** We first examined people's professed reasons for supporting/opposing trans-inclusive policies. To test whether opponents of trans-inclusive policies indicated that male violence was more relevant to their policy stance, while supporters indicated that their attitudes towards trans people was more relevant, we meta-analyzed (a) the interaction between opposition/support (opponent vs. supporter) and reason (attitudes towards trans people vs. male violence), and (b) their simple effects, separately for opponents and supporters. We converted effect sizes to Pearson's  $r$  to facilitate comparisons across analyses.

Opponents indicated that male violence was a more important reason than their attitudes towards trans people,  $r=.150$ ,  $SE=.038$ ,  $p<.001$ , 95% CI [.076, .225], while the opposite was the case for supporters,  $r=.237$ ,  $SE=.036$ ,  $p<.001$ , 95% CI [.166, .308], as predicted (see Figure 1). This interaction was significant,  $r=.188$ ,  $SE=.033$ ,  $p<.001$ , 95% CI [.123, .252], suggesting that opponents and supporters differ in which reason they voice as more important.

**Figure 1**

*Importance of Different Reasons Indicated by Opponents and Supporters of Trans-inclusive Policies (Studies 1-4)*



*Note.* Error bars refer to 95% confidence intervals.



**Actual predictors of policy support (Hypothesis H1b).** Did supporters' and opponents' professed reasons for policy support match the data? If people's professed reasons were accurate, then male violence beliefs should be more strongly associated with policy stances (among opponents), while trans attitudes should be more strongly associated with policy stances (among supporters). On the other hand, if – as we predicted – trans attitudes are driving policy stances, then trans attitudes should be more strongly associated with policy stances among both supporters *and* opponents. In other words, we tested whether people's self-endorsed reasons for policy support, as described above, were reflected in their own data by comparing whether the correlation between trans attitudes and policy stance was stronger than the corresponding correlation between gender-violence beliefs and policy stance.

To this end, we meta-analyzed the results from a series of preregistered Williams' *t*-tests comparing the correlation between gender-violence associations (versus trans attitudes) and support for trans-inclusive policies (both implicit and explicit), separately for supporters and opponents.

Supporters reported that trans attitudes influenced their policy stance more than gender violence beliefs. And, for supporters, both implicit trans attitudes:  $r = -.163$ ,  $SE = .030$ ,  $p < .001$ , 95% CI [-.221, -.105], and explicit trans attitudes:  $r = -.262$ ,  $SE = .028$ ,  $p < .001$ , 95% CI [-.318, -.207] predicted support for trans-inclusive policies, such that those who had more negative attitudes supported the policies less. Male violence beliefs were also correlated with support but less strongly and in the opposite direction,  $r = .134$ ,  $SE = .030$ ,  $p < .001$ , 95% CI [.075, .192]. In other words, supporters who associated men with violence were *more* supportive of trans-inclusive policies. The difference between the gender violence and trans attitude correlations was significant for both explicit and implicit trans attitudes,  $r = .213$ ,  $SE = .029$ ,  $p < .001$ , 95% CI [.156, .270] and  $r = .277$ ,  $SE = .028$ ,  $p < .001$ , 95% CI [.222, .332]

respectively. That is, in line with what supporters themselves reported, trans attitudes predicted their policy support more strongly than did male violence beliefs.

Opponents reported that gender violence beliefs influenced their policy stance more than trans attitudes. And, for opponents, the association of men with violence was linked to reduced support, such that those people who more strongly associated men with violence were *less* supportive of trans-inclusive policies,  $r = -.136$ ,  $SE = .051$ ,  $p = .007$ , 95% CI [-.235, -.036]. However, explicit trans attitudes also predicted support for trans-inclusive policies and did so more strongly, such that those with more negative attitudes supported policies less,  $r = -.328$ ,  $SE = .044$ ,  $p < .001$ , 95% CI [-.414, -.242]. Implicit trans attitudes, however, were unrelated,  $r = -.106$ ,  $SE = .093$ ,  $p = .254$ , 95% CI [-.288, .076]. That is, as predicted (and contrary to what opponents themselves report), explicit trans attitudes were a stronger predictor of policy support than associations with male violence,  $r = .146$ ,  $SE = .053$ ,  $p = .006$ , 95% CI [.042, .251]. This difference was not significant for implicit attitudes,  $r = -.015$ ,  $SE = .076$ ,  $p = .847$ , 95% CI [-.163, .134]. Taken together, these results show that, as predicted, opponents of trans-inclusive policies portray their reasons for support/opposition less accurately than supporters of these policies, at least when it comes to explicit trans attitudes.

## **Discussion**

We predicted that that opponents of trans-inclusive policies would portray their reasons for their policy stance less accurately than policy supporters. Across four studies, we found that supporters of trans-inclusive policies report (accurately) that their stance is most strongly predicted by their attitudes towards trans people. Opponents of trans-inclusive policies, on the other hand, claimed that their concerns about male violence were the primary reason for their opposition, but this was not reflected in their data. In other words, opponents of trans-inclusive policies do not accurately report what drives their opposition.

Why did opponents' self-reported reasons not match the data? One possibility is measurement mismatch. That is, the item measuring underlying reasons for policy views asks only about transgender people, while both the implicit and the explicit measures of trans attitudes (used as predictors) were relative measures of participants' preference of cisgender people over transgender people. Thus, if policy views are in reality mostly predicted by attitudes towards *cisgender* people (captured by the relative measures), our correlational comparison would be misleading. To address this possibility, in Study 5 we replaced the explicit measure with an absolute measure that exclusively focuses on transgender people, namely denying that transgender women are women.

A second possibility is that the gender-violence measure we used does not accurately reflect the arguments made by opponents of trans-inclusive policies. That is, their arguments may focus less on whether or not men are more violent than women (on average), but instead on the idea that a small minority of predatory men may abuse trans-inclusive policies to harm women, or prioritizing the overall importance of women's safety more generally. In Study 5, we therefore replaced the gender-violence measure with two new measures focusing on 1) the importance of women's safety and 2) predatory men.

### **Study 5: Conceptual Replication**

To rule out these possibilities, and test whether results generalize across different operationalizations of trans attitudes and gender-violence beliefs, we conducted a study using new measures of these predictors. Furthermore, rather than asking for causal reasons, we asked participants to report the perceived *association* between each predictor and their policy stance, to mirror our own empirical analyses.

## Method

### *Participants*

We aimed for a pre-registered ([https://aspredicted.org/1NX\\_XNL](https://aspredicted.org/1NX_XNL)) sample of 600 US American volunteers at Project Implicit and overshot slightly. Out of 652 participants, we excluded 23 participants for whom 10% of responses on the IAT are under 300 milliseconds, 16 who were transgender or non-binary or did not indicate their gender, 28 who did not self-categorize as opponents or supporters, and 20 under 18, resulting in a final sample size of 565, yielding 80% power to detect an effect size of  $f=.07$ .

### *Procedure and Measures*

Participants first completed two new measures relating to gender and violence, presented in random order, both with 7-point response scales from “strongly disagree” to “strongly agree.” Five items assessed the belief that predatory men pose a danger to women: “some men will use any opportunity they can to harass women,” “there are men who actively seek out opportunities to harm women,” “if there is some kind of loophole, some men will use it to harm women,” “there are men who are always on the lookout for new ways to hurt women,” and “certain men will take any chance they get to gain access to vulnerable women” ( $\alpha=.89$ ).

Five items assessed the importance of women’s safety: “women have a right to feel safe,” “women should not have to live in fear of male violence,” “it is important that women can feel safe from male violence,” “more needs to be done to ensure that women and girls are safe,” and “I don’t care very much about women’s safety” (reverse scored). To increase reliability, we excluded the last item ( $\alpha=.64$ ).<sup>2</sup> Both measures were specifically developed for this study.

---

<sup>2</sup> Results remain the same when the reverse-scored item is included.

Next, participants read a definition of “cisgender” and “transgender” and responded to a new measure of trans prejudice denying womanhood to transgender women (REFERENCE BLINDED FOR REVIEW): “Transgender women identifying as lesbians is problematic,” “womanhood is defined by biological sex,” “anyone who identifies as a woman is a woman,” (reverse-scored), and “transgender women can never truly understand what it means to grow up as a girl in a patriarchal society” ( $\alpha=.75$ ).

Participants then self-categorized as opponents or supporters of trans-inclusive policies, filled out the same measure of support for trans-inclusive policies as in Studies 3-4 ( $\alpha=.86$ ), and indicated the role of attitudes towards trans people and male violence in their levels of support. Finally, to more closely mirror our correlational analyses, we removed causal language (“reason”) from these items and changed the wording to “...were strongly related to my support for/opposition to the policies.” Lastly, participants took the same IAT as in the previous studies.

## **Results**

Descriptive statistics and correlations are displayed in Table 4.

Table 4

*Descriptive Statistics and Correlations for Opponents and Supporters of Trans-inclusive Policies (Study 5)*

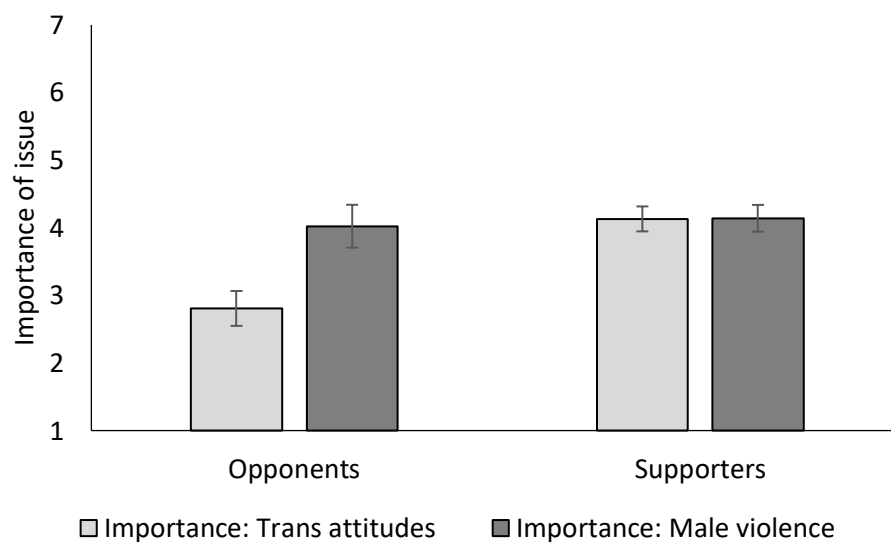
	Descriptive Statistics				Correlations						
	Opponents		Supporters		1	2	3	4	5	6	7
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>							
1. Trans attitudes IAT d score	0.31	0.37	0.14	0.41	-	.20***	-.05	.07	-.20***	-.07	-.08
2. Denial of womanhood	4.48	1.29	2.78	1.24	.08	-	-.14**	-.03	-.62***	-.27***	-.07
3. Concerns regarding predatory men	5.19	1.32	5.54	1.21	-.11	-.10	-	.34***	.10	.04	.14**
4. Importance of women's safety	6.49	0.91	6.73	0.66	.00	.06	.38***	-	.01	.02	.05
5. Support for trans-inclusive policies	3.10	1.19	4.94	1.00	-.05	-.54***	.11	.06	-	.29***	.08
6. Importance: Trans attitudes	2.81	1.72	4.13	1.83	.02	.09	-.11	-.16*	-.01	-	.34***
7. Importance: Male violence	4.02	2.12	4.14	1.94	-.02	-.05	.13	.12	.15*	.27***	-

*Note.* \*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .001$ . Correlations above the diagonal refer to those for supporters of trans-inclusive policies ( $n=387$ ) while those below the diagonal refer to correlations of opponents ( $n=178$ ).

**Perceived relationships with policy support (Hypothesis 1a).** Consistent with Studies 1-4, we predicted that opponents of trans-inclusive policies would once again indicate that their belief that men are violent was more related to their policy stance, while supporters would indicate that trans attitudes were more related. We tested this hypothesis in a 2 (Policy views: Supporter vs. Opponent) X 2 (Policy Relevance: Trans attitudes vs. Male violence) mixed ANOVA with perceived relatedness as a within-participants factor. We found a main effect of policy views,  $F(1, 541)=25.41, p<.001, \eta_p^2=.05$ , such that supporters ( $M_{ME}=4.13, SE=.08$ ) rated *both* issues as more related to their policy stance than opponents ( $M_{ME}=3.41, SE=.17$ ) and a main effect of relevance,  $F(1, 541)=35.29, p<.001, \eta_p^2=.06$ , such that overall male violence ( $M=4.09, SD=2.00$ ) was seen as more related to policy beliefs than trans attitudes ( $M=3.70, SD=1.80$ ).

**Figure 2**

*Importance of Different Issues Indicated by Opponents and Supporters of Trans-inclusive Policies (Study 5)*



*Note.* Error bars refer to 95% confidence intervals.

Replicating findings from Studies 1-4, however, these main effects were qualified by the predicted interaction,  $F(1, 541)=34.67, p<.001, \eta_p^2=.06$  (see Figure 2). That is,

opponents of trans-inclusive policies reported that concerns about male violence were more related to their policy stances,  $p < .001$ . Contrary to previous findings, supporters reported that trans attitudes and male violence were equally related,  $p = .963$ .

**Actual relationships with policy support (Hypothesis 1b).** Consistent with Studies 1-4, opponents' perceptions were not reflective of the patterns in their data (see Table 4). Indeed, neither concerns regarding predatory men nor the importance of women's safety were associated with support for trans-inclusive policies (among opponents) - but the denial of transgender women's womanhood was. We conducted a series of Williams' t-tests to formally compare the relative strength of these dependent correlations. Among opponents, explicit trans attitudes (in this case, the denial of womanhood) were more strongly related to policy support than concerns about predatory men,  $t(174) = -6.76$ ,  $p < .001$ , or the importance of women's safety,  $t(174) = -6.37$ ,  $p < .001$ . This difference was not significant for implicit trans attitudes (predatory men:  $t(174) = -1.45$ ,  $p = .148$ ; women's safety:  $t(174) = -1.02$ ,  $p = .309$ ).

Consistent with our meta-analytic results, among supporters the explicit measure of trans attitudes was once again more strongly related to policy stance than concerns about predatory men,  $t(381) = -11.40$ ,  $p < .001$ , or importance of women's safety,  $t(381) = -10.50$ ,  $p < .001$ . The same was true for implicit attitudes (predatory men:  $t(381) = -4.13$ ,  $p < .001$ ; women's safety:  $t(381) = -3.10$ ,  $p = .002$ ). That is, while supporters accurately predicted the importance of trans attitudes to their policy stance, they – like opponents – overestimated the relative importance of violence concerns.

## Discussion

Using a new set of measures to operationalize male violence concerns and trans attitudes, we largely replicated meta-analytic results from Studies 1-4. Furthermore, we did so even when asking participants to merely predict which of the two reasons – male violence or trans attitudes – was more related to their own policy stances.



Consistent with studies 1-4, opponents predicted male violence concerns were more strongly related to their policy views than trans attitudes, but this was not reflected in their data, which showed trans attitudes to be a stronger predictor. Surprisingly, in this study, supporters of trans-inclusive policies predicted that trans attitudes were related to their policy stances (as in Studies 1-4), but (unlike Studies 1-4) also predicted male violence concerns would be equally related. Because we adjusted the wording of this measure (to ask about relationships rather than causes) and first had participants complete the new measures (rather than counterbalancing), it is difficult to say why this was the case. Thus, overall, while supporters were accurate in reporting that trans attitudes were related to their policy stances, they inaccurately overestimated the importance of male violence beliefs to their policy views.

Taken together, Studies 1-5 demonstrate that while opponents of trans-inclusive policies claim that their opposition is primarily based on concerns about male violence and women's safety, this is not reflected in their data: Opposition is more strongly predicted by explicit trans attitudes compared to male violence concerns. This effect replicates across multiple operationalizations of trans attitudes, trans policy beliefs, male violence, and women's safety, and is robust to whether participants are asked to report on the causes (versus correlates) of their policy stances.

However, because the studies thus far are correlational, they have several limitations. Perhaps most importantly, we cannot make causal claims about the direction of the observed effects; it is possible that policy stances drive trans attitudes (rather than vice versa), or that third variables account for both. While we tried to address this problem by asking participants to predict the relative strength of associations with their policy stances (rather than reasons) in Study 5, it could nevertheless be the case that opponents in Studies 1-4 *did* accurately report their reasons - and that the stronger correlation between trans attitudes and policy opposition was due to other causes. In addition, because the observed correlational data

measure associations at the group-level, whereas participants are citing reasons at the individual-level, we cannot rule out the possibility that our results reflect these different levels of analysis rather than accuracy per se.

To address these limitations, we conducted two experiments to test whether changing people's trans attitudes and beliefs about male violence would produce corresponding shifts in trans-inclusive policy support (or lack thereof). If manipulating trans attitudes (but not male violence beliefs) changes policy support, this would strengthen our claim that opponents of trans-inclusive policies do not report the true reasons for their support. However, we note that providing such causal evidence poses several challenges: First, it is notoriously difficult to reliably manipulate attitudes and/or stereotypes, even temporarily. One of the most robust published manipulations of trans prejudice involved extensive in-depth in-person conversations - and only reduced trans prejudice by  $d=.08$  (Kalla & Broockman, 2020). And, ideally, such a manipulation would manipulate both trans attitudes and male violence stereotypes separately in parallel (i.e., to eliminate confounds) and with similar manipulation strength, to allow for direct comparisons. With these limitations in mind, we sought to manipulate transgender attitudes and male-violence beliefs in Studies 6-7.

### **Study 6: Manipulating Trans Attitudes**

In Study 6, we attempted to experimentally manipulate trans attitudes and male violence beliefs using an adapted version of the *imagined contact paradigm*. Positive intergroup contact can reduce negative attitudes towards outgroups, including the elderly, Muslims, homeless people, and – most relevant to our research – transgender people (West et al., 2017). Similar results can be obtained by asking people to merely *imagine* having a positive interaction with an outgroup member and write down the positive things they are imagining.

While this paradigm has, to our knowledge, not been used to manipulate stereotypes, we adapted it here to manipulate male violence beliefs in an effort to have manipulations that were as parallel as possible. We randomly assigned participants to imagine a positive interaction with a person identified as either a trans woman (positive trans condition), a kind soft-spoken man (peaceful man condition), or (in the control condition) no additional information. We predicted that the positive trans condition – but not the peaceful man condition – would increase subsequent trans-inclusive policy support. To maximize statistical power, we included only opponents of trans-inclusive policies. Because Studies 1-5 clearly established that opponents of trans-inclusive policies report male violence concerns (versus trans attitudes) to be the more important to their policy views, we did not ask about reasons in this study to reduce participant burden.

## **Method**

### ***Participants***

As preregistered, ([https://aspredicted.org/85V\\_PKB](https://aspredicted.org/85V_PKB)), we recruited 700 US American participants on Prolific and excluded two participants who identified as transgender, resulting in a final sample size of 698, yielding 80% power to detect an effect size as small as  $f=.12$ . We used a pre-screening survey to identify opponents of trans-inclusive policies.

### ***Procedure and Materials***

We recruited US American participants across the political spectrum (1/3 liberal, 1/3 moderate, 1/3 conservative) for a pre-screening survey to identify opponents of trans-inclusive policies for the main study. Participants indicated whether, in general they were opposed to or supportive of trans-inclusive policies, using the same wording as in previous studies. To hide the purpose of the main study, we also asked about two unrelated topics (their views on policies regarding COVID and sex work). Opponents of trans-inclusive

policies then had the opportunity to access the main study, advertised as a study on how people imagine different scenarios, two weeks later.

Participants were randomly assigned to one of three conditions: an adapted imagined contact paradigm (West et al., 2017) aimed to improve attitudes towards trans people (positive trans condition), aimed to decrease male violence beliefs (peaceful man condition), or a control condition. Participants in all conditions imagined attending a friend's birthday celebration at a restaurant and wrote in detail about the experience for five minutes. In the positive trans condition, they were told to imagine sitting opposite a trans woman and having a positive interaction with her. In the male peacefulness condition, they were asked to imagine sitting opposite and interacting with a man who was kind, soft-spoken, and gentle.

After completing the imagined contact task, participants indicated their support for trans-inclusive policies ( $\alpha=.88$ ) as well as their trans attitudes ( $\alpha=.88$ ) and gender-violence beliefs ( $\alpha=.88$ ) using the same measures as in Study 4. They then provided demographic information and were debriefed.

## Results

Descriptive statistics and correlations are displayed in Table 5. Note that in this study, beliefs about male violence were not associated with policy views, even though all participants were opposed to trans-inclusive policies.

**Table 5**  
*Descriptive Statistics and Bivariate Correlations (Study 6)*

	Overall Sample		Positive trans condition		Peaceful man condition		Control condition		Correlations	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	2	3
1. Trans attitudes	4.93	0.99	4.85	0.93	4.85	0.99	5.07	1.02	.11**	-.45***
2. Gender-violence association	5.05	0.84	5.07	0.85	5.01	0.83	5.07	0.83		.04
3. Support for trans-inclusive policies	3.08	1.62	3.32	1.63	2.96	1.54	2.96	1.65		

*Note.* \*\*  $p < .01$ ; \*\*\*  $p < .001$

### *Manipulation Checks*

We first checked whether our manipulation was successful in shifting trans attitudes and gender-violence beliefs using two one-way ANOVAs with condition as the independent variable. The effect of condition on trans attitudes was significant,  $F(2, 695)=4.14, p=.016, \eta_p^2=.01$ . Post-hoc tests revealed that, as intended, trans attitudes in the positive trans condition were less negative compared to the control condition,  $p=.032$ . Unexpectedly, trans attitudes in the peaceful man condition also reduced negative trans attitudes compared to the control condition,  $p=.042$ , and did so to the same extent as the transwoman condition,  $p=.996$ . In other words, both experimental conditions decreased negative attitudes towards transgender people.. Condition had no effect on the gender-violence association,  $F(2, 695)=0.38, p=.684, \eta_p^2<.01$ .

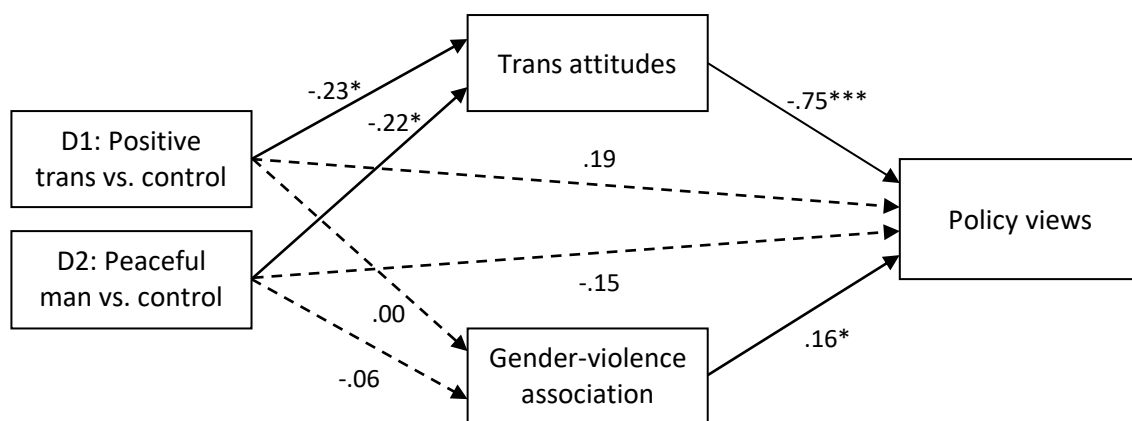
### *Pre-registered analysis*

Did shifting trans attitudes affect policy views? Condition did affect policy views,  $F(2, 695)=3.79, p=.023, \eta_p^2=.01$  and post-hoc tests revealed that, as predicted, support for trans-inclusive policies was higher in the positive trans condition compared to the control condition,  $p=.041$ , and compared to the peaceful man condition,  $p=.049$ . The latter two conditions did not differ from each other,  $p>.999$ .

*Exploratory analyses*

Because the peaceful man condition unexpectedly shifted trans attitudes, but this shift was not reflected in policy views, we used PROCESS (Hayes, 2018; v4.0, Model 4) to test for indirect effects of the different experimental conditions through trans attitudes and gender-violence beliefs, respectively. We dummy-coded condition with the control condition as the reference category and entered gender-violence beliefs and trans attitudes as parallel mediators (see Figure 3).

**Figure 3**  
*Mediation Model Predicting Policy views*



*Note.* Higher values in trans attitudes indicate more negative attitudes; higher value in the gender-violence association indicate a stronger association of men with violence; higher values in policy views indicate more supportive views of trans-inclusive policies.

Dashed lines indicate non-significant associations.

\*  $p < .05$ ; \*\*\*  $p < .001$

Trans attitudes significantly mediated the effects of both the positive trans condition,  $B = .17$  [.04, .31], and the peaceful man condition,  $B = .16$  [.03, .30]. That is, to the extent that the condition reduced negative trans attitudes, it also reduced opposition to trans-inclusive policies. In contrast, gender-violence beliefs did not mediate the effects of either condition (positive trans condition:  $B = .0003$  [-.03, .03]; peaceful man condition:  $B = -.01$  [-.04, .02]. Taken together, this suggests that the peaceful man condition was (inadvertently) successful

at manipulating trans attitudes, and to the extent it did so, it also reduced policy opposition. However, the condition appears to have also manipulated other unmeasured constructs that masked this effect.

## **Discussion**

Our primary goal was to provide causal evidence showing that trans attitudes, compared to male violence beliefs, *cause* opposition to trans-inclusive policies. As predicted, participants who imagined a positive interaction with a trans woman had less negative attitudes towards trans people compared to the control condition. This difference was also reflected in policy views. Unexpectedly, imagining interacting with a kind and gentle man *also* reduced negative attitudes towards trans people. While this was not our intention, an exploratory mediation analysis showed that, as we would expect, this change in trans attitudes was also associated with less opposition to trans-inclusive policies.

In summary, this study provided causal evidence that attitudes towards trans people affect support for trans-inclusive policies. However, we were unable to compare the strength of this causal effect with that of male violence beliefs due to the ineffective manipulation of male violence beliefs. This is perhaps not surprising given that the imagined contact paradigm has historically been used to shift attitudes, and not group stereotypes. In our next study, we therefore adapted a manipulation that has been used to manipulate stereotypes: imagined futures.

### **Study 7: Manipulating Gender-Violence Beliefs**

In Study 6, we were able to examine the causal effect of shifting trans attitudes, and found that changing attitudes led to increased policy support. In Study 7, we attempted to do the same for male violence beliefs, using a stereotype change paradigm developed by Koenig and Eagly (2014). In this approach, participants picture a future in which certain social groups (e.g., women versus men) are over-represented in roles associated with specific

attributes (e.g., agency and communion). If women are described as being particularly well-represented in groups associated with agency (e.g., in leadership positions), women are then seen as possessing more agentic qualities in this imagined future society.

We adapted this paradigm by randomly assigning participants to picture a future society where transgender (versus cisgender) people are well-represented in highly liked (versus disliked) occupations, and where men (versus women) are overrepresented in peaceful (versus violent) occupations. Once again, we predicted that trans attitudes, compared to views about male violence, would more strongly affect policy views, counter to what opponents reported in Studies 1-5.

## **Method**

### ***Participants***

We recruited a preregistered sample of 920 US American participants on Prolific ([https://aspredicted.org/W22\\_Y8Z](https://aspredicted.org/W22_Y8Z))<sup>3</sup>, using the same recruitment strategy as Study 6. We excluded nine participants who withdrew their consent after being debriefed, six transgender participants, and one participant under the age of 18. A total of 176 participants failed our attention check, whom we excluded in keeping with our pre-registration (primary conclusions do not change when included). The final sample size was thus 728, giving us 80% power to detect an effect size as small as  $f=.10$ .

### ***Procedure and Measures***

Participants were randomly assigned to one out of four conditions in a 2 (Trans attitude: Positive vs. Negative) X 2 (Gender-violence: Peaceful men vs. Violent men) between- participants design. Participants were first told about predicted changes in what US society would look like in 25-30 years, adapted from Koenig and Eagly (2014). In the positive trans attitudes condition, they were told that transgender people would be

---

<sup>3</sup> One of the pre-registered analyses is reported in the online supplement.



particularly well-represented in highly-liked occupations such as firefighter, nurse, and among people working for pet rescue organizations, while cisgender people would be particularly well-represented among telemarketers, car salespeople, and lawyers (i.e., disliked occupations). In the negative trans attitudes condition, the occupations were flipped.

Moreover, participants were told about which occupations men and women would be particularly well-represented in. In the peaceful men condition, they were told that men would be particularly well-represented in peaceful occupations such as yoga instructor, massage therapist, and florists, while women would be particularly well-represented in violent occupations such as police officer, bouncer, and athletes (such as hockey players and boxers). In the violent men condition, the occupations were flipped. Roles were selected based on a pilot-study (see online supplement).

Participants were asked to assume that these predictions were correct and to imagine living in this future society. They then indicated the extent to which they would support a range of different policies in this future society. The policies included filler items (e.g., “In this future society, more domestic violence resources need to be made available to men”) as well as adapted versions of the items used to measure support for trans-inclusive policies in Studies 1-6 (e.g. “In this future society, transgender women should have access to women-only support groups for victims of sexual or domestic abuse.”  $\alpha=.88$ ).

On the next two pages, they responded to items measuring their gender-violence association ( $\alpha=.91$ ) and their trans attitudes ( $\alpha=.90$ ) in randomized order, using the same items (adapted to refer to the imagined future society) as in Study 6. Participants then responded to attention check items, asking them to recall which groups transgender people and women were predicted to be well-represented in.

## Results

Descriptive statistics and correlations are displayed in Table 6. Note that, once again, beliefs about male violence were not associated with policy views.

**Table 6**  
*Descriptive Statistics and Bivariate Correlations (Study 6)*

	<i>M</i>	<i>SD</i>	Correlations	
			2	3
1. Trans attitudes	5.01	1.03	-.08*	-.51***
2. Gender-violence association	4.62	0.99		-.01
3. Support for trans-inclusive policies	3.33	1.63		

*Note.* \*\*  $p < .01$ ; \*\*\*  $p < .001$

### *Manipulation Checks*

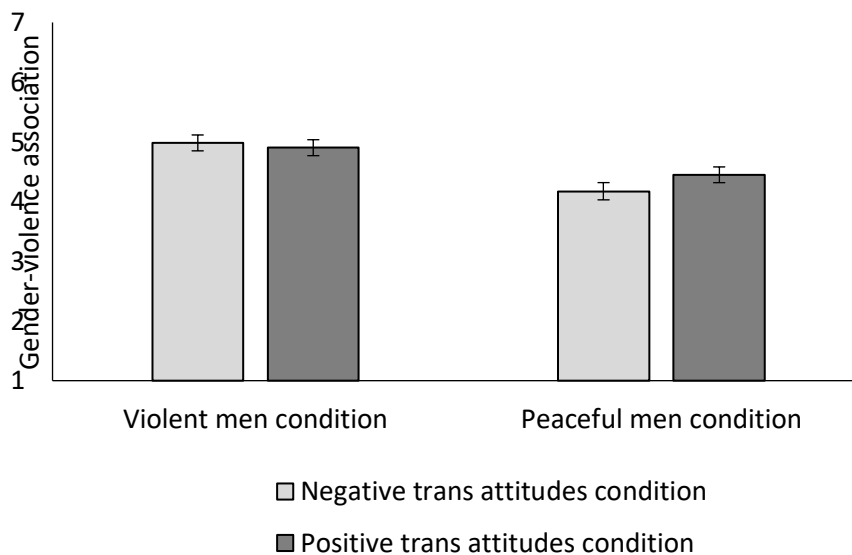
We first checked whether our manipulation was successful in shifting trans attitudes and the gender-violence association using two 2(Trans attitude: Positive vs. Negative) X 2 (Gender-violence: Peaceful men vs. Violent men) ANOVAs.

For male violence beliefs, we found a main effect of the peaceful man condition,  $F(1,724)=83.68$ ,  $p < .001$ ,  $\eta_p^2=.10$ , such that the association of men with violence was lower in the peaceful men condition ( $M=4.31$ ,  $SD=0.97$ ) compared to the violent men condition ( $M=4.94$ ,  $SD=0.90$ ). The main effect of the positive trans condition on male violence beliefs was not significant,  $F(1,724)=2.00$ ,  $p=.158$ ,  $\eta_p^2 < .01$ . Unexpectedly, the interaction between the two conditions was significant,  $F(1,724)=6.61$ ,  $p=.010$ ,  $\eta_p^2=.0$ . As expected, the gender-violence manipulation was successful in both the negative trans attitudes condition,  $p < .001$ , and the positive trans attitudes condition  $p < .001$ . However, in the peaceful men condition, there was an unexpected effect of the trans attitudes condition,  $p=.004$ , such that the association of men with violence was lower in the positive trans attitudes condition (see Figure 4). In other words, the gender violence manipulation had a stronger effect when transgender people were over-represented in occupations that people feel more positive towards.

For trans attitudes, none of the effects were significant (all  $F_s < 2.79$ , all  $p_s > .105$ ), indicating that our manipulation did not successfully shift attitudes towards transgender people.

**Figure 4**

*Gender-Violence Association Based on Condition (Study 7)*



*Note.* Error bars refer to 95% confidence intervals.

### ***Pre-registered Analysis***

We ran a 2 (Trans attitude: Positive vs. Negative) X 2 (Gender-violence: Peaceful men vs. Violent men) ANOVA to test the effect of both manipulations on support for trans-inclusive policies. If opponents of trans-inclusive policies accurately report their reasons for opposition, the observed difference in male violence belief should be reflected in a main effect of the gender violence manipulation. That was not the case,  $F(1,724)=2.03$ ,  $p=.155$ ,  $\eta_p^2 < .01$ . Given that the manipulation of trans attitudes was unsuccessful, we did not expect to see a main effect of the positive trans condition on policy support, and we did not,  $F(1,724)=0.02$ ,  $p=.895$ ,  $\eta_p^2 < .01$ .

Unexpectedly, the interaction between the positive trans condition and peaceful man condition was once again significant,  $F(1,724)=6.29$ ,  $p=.012$ ,  $\eta_p^2=.01$ . Namely, in the negative trans attitudes condition, support for trans-inclusive policies was lower in the violent men condition than in the peaceful men condition,  $p=.005$  (see Figure 5). None of the other differences were significant.

**Figure 5**

*Support for Trans-inclusive Policies Based on Condition (Study 7)*



*Note.* Error bars refer to 95% confidence intervals.

## Discussion

In Study 7, we successfully manipulated male violence beliefs, such that participants in the peaceful man condition reported lower gender-violence associations than in the violent man condition. However, despite this, participants who imagined living in a future society where men were more peaceful were not in turn more supportive of trans-inclusive policies than those who imagined living in a society in which men were violent. Thus, this study contradicts the claim made by opponents of trans-inclusive policies that cites concerns about male violence as the reason for their opposition. Lowering gender-violence associations had no impact on policy support, and indeed, envisioning a society where men were indeed less

violent did not affect policy support – as would be expected if it were the true underlying cause for opposing such policies.

Unfortunately, we were not able to successfully manipulate transgender attitudes in this study, and were thus unable to compare the relative effects of shifting trans attitudes versus gender-violence beliefs directly. We believe this underscores the difficulty of finding a manipulation that is effective in pushing around both attitudes and stereotypes of separate groups, and doing so with comparable effect sizes.

On this note, we observed an unexpected interaction of our trans attitude manipulation with the gender violence manipulation on both gender-violence beliefs and policy views (but *not* trans attitudes). In short, the gender-violence manipulation was more effective in shifting beliefs about male violence when picturing a society where transgender people were over-represented in disliked occupations. Furthermore, in this condition, the peaceful man condition *did* increase policy support, suggesting that, at least under some circumstances, beliefs about male violence may affect policy views. However, because the manipulation check for trans attitudes indicated that we were *not* manipulating attitudes, it is unclear what exactly these circumstances might be. Given that this interaction effect was not predicted and quite small, it needs to be replicated and should be interpreted with caution.

### **General Discussion**

Opponents of trans-inclusive policies, particularly policies that grant transgender women access to women-only spaces, often argue that they oppose such policies not because they hold negative attitudes towards trans people but because they are concerned that such policies will give predatory men access to these spaces and, in turn, the opportunity to harm women. Across seven studies using a wide array of operationalizations, we provide correlational and experimental evidence that these arguments do not reflect reality: Counter to what opponents themselves reported, opposition to trans-inclusive policies was

consistently more strongly predicted by (explicit) trans attitudes compared to male violence concerns. Moreover, in Studies 6-7, manipulating explicit trans attitudes – but not male violence associations – created shifts in policy support.

These findings are in line with psychological research showing that people can often have little insight into the reasons underlying their behaviors and beliefs (e.g., Nisbett & Wilson, 1977; Wilson, 2004), and may be viewed as an instance of *covering* (Crandall & Eshleman, 2003), that is, obscuring one's prejudicial views by providing an alternative explanation. Indeed, for supporters of trans-inclusive policies, concerns about male violence were, if anything, associated with *higher* levels of support. This may reflect beliefs about who needs to be protected from such violence and the awareness that transgender women in men-only spaces are at a higher risk for becoming victims of male violence.

### **Theoretical and Practical Implications**

This work has important theoretical and practical implications for understanding the causes and consequences of anti-transgender prejudice. In 2022 alone, over 150 anti-transgender bills have been proposed in the United States – more than in any previous year (ACLU, 2022). These bills adversely affect transgender people, for example by forcing transgender people to use facilities associated with their sex assigned at birth rather than their gender identity or by making it more difficult to change their gender on legal documents. Similarly, in the UK, a court ruled in 2021 that children under the age of 16 could no longer be prescribed puberty-blocking drugs (although this decision was later overturned on appeal; Siddique, 2021). Such legislation can have dramatic consequences on the mental health of those affected by them, up to and including suicide (e.g., Rew et al., 2021). Given the rising numbers of individuals who openly identify as transgender, it is important to understand and find ways to attenuate the psychological forces driving this surge in anti-trans legislation.

Our findings provide an important first step by showing that the reasons cited by opponents of such policies should not necessarily be taken at face value. Improving the safety of women, for instance, is unlikely to translate into greater support for trans-inclusive policies, given that we find no empirical evidence that safety concerns are tied to reduced policy support. Relying on the overt discourse surrounding trans-inclusive policies and taking professed reasons at face value - instead of empirically examining the underlying psychological mechanisms behind such beliefs – is thus likely to result in false conclusions about what drives legislative discrimination against transgender people. This is important not only practically – for countering such legislation – but also for better understanding the theoretical drivers of anti-trans discrimination, more broadly.

For instance, it is notable that opponents of trans-inclusive policies do not attribute their policy stances to their feelings about transgender people, despite the correlational and experimental evidence reported here showing that they are indeed very much related. This is, however, consistent with a long history in the literature documenting reluctance to openly disclose one’s own biases or prejudice.

Likewise, our hypothesis that opponents would not accurately report their reasons were less consistent for UK samples (Studies 3-4) than for US samples (Studies 1, 2, and 5; see Tables 1-2 and online supplement). This is perhaps unsurprising. Values of equality and equal opportunity are a key American value (see O’Brien et al., 2010). Thus, Americans may be more motivated to appear non-prejudiced, both to themselves and to others, while British people may feel more comfortable openly acknowledging that their opposition to trans-inclusive policies is driven by their negative attitudes towards trans people.

These results also illustrate the methodological limitations of asking people to explain the causes of their own behavior. Although long noted in psychology that people do not have privileged access to their own mental processes (e.g., Nisbett & Wilson, 1977) and thus

cannot accurately introspect on the cognitive processes driving their own behavior, asking people to report *why* they support a particular candidate, party, or policy is still commonplace in studying political behavior. While asking people to self-report their reasons is far easier than testing those causes empirically (as illustrated by this paper), accepting such answers at face value – rather than understanding them for the lay theories they are – hinders our ability to draw accurate scientific conclusions.

For instance, while we focused on policies that would grant transgender people access to gender/sex-segregated spaces, where women’s safety is primarily invoked, these are not the only policies for which trans-inclusion is relevant. Issues around fairness and gender equality are often mentioned in the context of transgender women’s access to women’s sports teams and of affirmative action to address gender inequality (see BBC, 2018; Magowan, 2018). Future research should investigate to what extent fairness perceptions are indeed at the core of opposition to such policies or if, once more, calls to “protect women” (in this case, women’s advancement) are merely used to disguise prejudice.

Finally, our findings add to the emerging literature on how beliefs about the role of women and men in society (e.g., feminist versus traditional views) intersect with beliefs about the nature of gender and sex more broadly (i.e., who counts as a woman or a man). Understanding this ideological intersection is particularly relevant for understanding the growing gender-critical feminist movement, which has been outspoken in opposing trans-inclusive policies. Gender-critical feminists, often referred to by the more controversial term TERFs (trans-exclusionary radical feminists), hold biology-based views of womanhood and object to identity-based concepts of gender. Importantly, in a recent quasi-representative study of US and UK feminists, gender critical views are not “fringe” – indeed, roughly half of self-identified feminists hold gender-critical view, with trans inclusion the most divisive issue among feminists (BLINDED FOR PEER REVIEW). While we did not investigate opposition



to trans-inclusive policies among feminists specifically, our findings nevertheless casts doubt on the claims made by prominent feminists such as J. K. Rowling, quoted at the beginning of this article.

Across five studies, we find consistently that opponents of trans-inclusive policies cite women's safety as driving their opposition. If this claim does not originate in people's actual causal reasoning, where does it come from and why is it so widely endorsed by opponents of such policies? The argument that (cis)women are in need of protection from violent men who seek to do them harm echoes benevolent sexist views that women should be cherished and protected (Glick & Fiske, 1996). Benevolent sexism is more widely accepted than other forms of sexism (Barreto & Ellemers, 2005) but associated with a range of outcome that *decreases* women's safety, such as sexual harassment (Fiske & Glick, 1995) and negative reactions to rape victims (Viki & Abrams, 2002).

Benevolent sexism works in tandem with hostile sexism (i.e., ambivalent sexism) to uphold the current gender system by seemingly rewarding women who adhere to gender norms and punishing women who go against such norms, such as women in leadership positions or feminists (Glick & Fiske, 1996). We suggest transgender women are another such group that is targeted for violating conventional cultural norms when it comes to gender, and that the concern about women's safety cited by policy opponents stems from this. For instance, safety concerns often focus on the perceived need to protect *cisgender* women, rather than transgender women. In one example, supporters of "bathroom bills" cite concern about the safety of cisgender women if transgender women are permitted to use women's bathrooms, but overlook the safety of transgender women forced (by the same policy) to use men's bathrooms. Notably, transgender women are a group not discussed or considered in the original conception of ambivalent sexism, suggesting that this theory may apply even in

novel, evolving social contexts. Future research should investigate the role of benevolent sexism in opposition to trans-inclusive policy stances more directly.

One means of extending this work is through a greater focus on transgender men. While our studies focused on transgender women because the current discourse focuses predominantly on transwomen, it would be interesting to examine whether similar arguments are used in the context of inclusion of transgender men. Interestingly, such arguments often also seem to be focused on protecting “women” and “girls” – in this case, lesbians and tomboys who are allegedly “brainwashed” into believing they are transgender and encouraged to do “irreversible damage” to their bodies (see Stahl, 2021).

### **Limitations and Future Research**

Our causal model and assumption, throughout the paper, is that trans attitudes are changing policy views, rather than vice versa. However, because Studies 1-4 are correlational, they cannot rule out the possibility that policy views are affecting transgender attitudes, or that third variables account for both. We addressed this possibility in two ways: first, in Study 5, rather than asking participants to report on the reasons for their views (for which we could not provide causal evidence), we instead asked them merely to predict the correlational strength of the relationship. That is, we compared the predicted strength of the relationship to the actual observed strength of the relationship, and found the same thing: trans attitudes are more strongly related to policy stances than male-violence beliefs. Second, and more importantly, in studies 6 and 7 we experimentally manipulated trans attitudes and male violence beliefs directly. While we observed increased policy support as a consequence of manipulating trans attitudes, we found no such effect of manipulating male violence beliefs, lending further experimental evidence that the causal chain flows from attitudes to policy, rather than vice versa.

Unfortunately, Studies 6 and 7 themselves had limitations, and we were not able to manipulate both constructs in the same study, as would be ideal for comparing the relative strength of both potential causes. This is, however, perhaps not surprising given that both stereotypes and attitudes are difficult to manipulate (in general) and even harder to change in parallel manipulations that eliminate confounding variables. What was effective in manipulating attitudes was not effective for manipulating stereotypes, and vice versa. Furthermore, the manipulations were quite different from each other. It is possible that something about these specific manipulations (imagined contact vs. imagining living in a future society) made it more or less likely for changes in stereotypes and attitudes to translate into policy views. Moreover, our manipulations were not as straightforward as we would have desired. In Study 6, the male violence manipulation did not affect male violence beliefs – but did affect trans attitudes, and in Study 7, the trans attitudes manipulation moderated the effect of our male violence manipulation for reasons that are unclear (since the trans attitude manipulation itself was unsuccessful at manipulating attitudes).

Nevertheless, we believe that, taken together, the findings from our experimental studies strengthen our claim. In Study 6, we successfully manipulated trans attitudes and these changes were reflected in policy views. In Study 7, we successfully manipulated male violence beliefs but these changes were *not* reflected in policy views. Future research should replicate and (hopefully) extend these findings by manipulating both simultaneously.

### **Conclusion**

Trans-inclusive policies are controversial, and opponents often claim that while they are supportive of trans people that cis women's safety needs to be protected. We find no evidence that concerns about male violence are the strongest predictor of such opposition; instead, negative attitudes towards transgender people are most strongly associated with opposition. Our findings have important implications for those campaigning for trans

inclusion, suggesting that the most effective strategies might be those aiming at changing attitudes rather than refuting arguments about the danger that trans inclusion allegedly poses to the safety of cisgender women.

## References

- Abramowitz, A. I. (1995). It's abortion, stupid: Policy voting in the 1992 presidential election. *The Journal of Politics*, *57*, 176-186.
- American Civil Liberties Union (2022, July 1). *Legislation affecting LGBTQ rights across the country*. <https://www.aclu.org/legislation-affecting-lgbtq-rights-across-country>
- Axt, J., Conway, M., Westgate, E., & Buttrick, N. (2020). Implicit transgender attitudes independently predict beliefs about gender and transgender people. *Personality and Social Psychology Bulletin*. Advance online publication.
- Barreto, M., & Ellemers, N. (2005). The burden of benevolent sexism: How it contributes to the maintenance of gender inequalities. *European Journal of Social Psychology*, *35*(5), 633-642.
- BBC (2018, May 1). *Labour: Row over inclusion of trans women in all-women shortlists*. Retrieved from <https://www.bbc.co.uk/news/uk-politics-43962349>
- Crandall, C. S., & Eshleman, A. (2003). A justification-suppression model of the expression and experience of prejudice. *Psychological Bulletin*, *129*(3), 414–446.  
<https://doi.org/10.1037/0033-2909.129.3.414>
- Crandall, C. S., Eshleman, A., & O'Brien, L. T. (2002). Social norms and the expression and suppression of prejudice: The struggle for internalization. *Journal of Personality and Social Psychology*, *82*, 359–378.
- Fernbach, P. M., Hagemayer, Y., & Sloman, S. A. (2014). Effort denial in self-deception. *Organizational Behavior and Human Decision Processes*, *123*(1), 1-8.
- Fiske, S. T., & Glick, P. (1995). Ambivalence and stereotypes cause sexual harassment: A theory with implications for organizational change. *Journal of Social Issues*, *51*, 97–115

- Glick, P., & Fiske, S. T. (1996). The ambivalent sexism inventory: Differentiating hostile and benevolent sexism. *Journal of Personality and Social Psychology*, *70*, 491–512
- Goh, J. X., Hall, J. A., & Rosenthal, R. (2016). Mini meta-analysis of your own studies: Some arguments on why and a primer on how. *Social and Personality Psychology Compass*, *10*(10), 535–549.
- Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: An improved scoring algorithm. *Journal of Personality and Social Psychology*, *85*(2), 197-216. <http://doi.org/10.1037/0022-3514.85.2.197>
- Hasenbush, A., Flores, A. R., & Herman, J. L. (2019). Gender identity nondiscrimination laws in public accommodations: A review of evidence regarding safety and privacy in public restrooms, locker rooms, and changing rooms. *Sexuality Research and Social Policy*, *16*(1), 70-83.
- Hayes, A. F. (2018). *Introduction to Mediation, Moderation, and Conditional Process Analysis, Second Edition: A Regression-Based Approach*. Guilford Press
- Howell, J. L., Redford, L., Pogge, G., & Ratliff, K. A. (2017). Defensive responding to IAT feedback. *Social Cognition*, *35*, 520-562.
- Kalla, J. L., & Broockman, D. E. (2020). Reducing exclusionary attitudes through interpersonal conversation: Evidence from three field experiments. *American Political Science Review*, *114*(2), 410-425
- Kirby, T. A., Tabak, J. A., Ilac, M., & Cheryan, S. (2020). The symbolic value of ethnic spaces. *Social Psychological and Personality Science*, *11*(7), 867-878.
- Kreps, T. A., Laurin, K., & Merritt, A. C. (2017). Hypocritical flip-flop, or courageous evolution? When leaders change their moral minds. *Journal of Personality and Social Psychology*, *113*(5), 730.

- Magowan, A. (2018, December 18). *Transgender women in sport: Are they really a 'threat' to female sport?* Retrieved from <https://www.bbc.co.uk/sport/46453958>
- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: verbal reports on mental processes. *Psychological Review*, *84*(3), 231-259.
- NISVS (2010). *National Intimate Partner and Sexual Violence Survey*. Retrieved from [https://www.cdc.gov/violenceprevention/pdf/NISVS\\_Report2010-a.pdf](https://www.cdc.gov/violenceprevention/pdf/NISVS_Report2010-a.pdf)
- O'Brien, L. T., Crandall, C. S., Horstman-Reser, A., Warner, R., Alsbrooks, A., & Blodorn, A. (2010). But I'm no bigot: How prejudiced White Americans maintain unprejudiced self-images. *Journal of Applied Social Psychology*, *40*(4), 917-946.
- Office of National Statistics (2017). *Analyses from the Crime Survey for England and Wales (CSEW) of sexual assaults experienced by adults aged 16 to 59, including by type, sex and personal characteristics, and police recorded sexual offences*. [Data set]. <https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/datasets/sexualoffencesappendixables>
- Paulhus, D. L., & Buckels, E. (2012). Classic self-deception revisited. In S. Vazire & T. D. Wilson (Eds.), *Handbook of self-knowledge* (p. 363–378).
- Rew, L., Young, C. C., Monge, M., & Bogucka, R. (2021). Puberty blockers for transgender and gender diverse youth—a critical review of the literature. *Child and Adolescent Mental Health*, *26*(1), 3-14.
- Rowling, J. K. (2020, June 10). *J.K. Rowling Writes about Her Reasons for Speaking out on Sex and Gender Issues*. Retrieved from <https://www.jkrowling.com/opinions/j-k-rowling-writes-about-her-reasons-for-speaking-out-on-sex-and-gender-issues/>
- Siddique, H. (2021, September 17). Appeal court overturns UK puberty blockers ruling for under-16s. *The Guardian*. <https://www.theguardian.com/society/2021/sep/17/appeal-court-overturns-uk-puberty-blockers-ruling-for-under-16s-tavistock-keira-bell>

Stahl, L. (2021, March 19). The latest form of transphobia: Saying lesbians are going extinct.

*The Washington Post*. Retrieved from [https://www.washingtonpost.com/outlook/the-latest-form-of-transphobia-saying-lesbians-are-going-extinct/2021/03/18/072a95fc-8786-11eb-82bc-e58213caa38e\\_story.html](https://www.washingtonpost.com/outlook/the-latest-form-of-transphobia-saying-lesbians-are-going-extinct/2021/03/18/072a95fc-8786-11eb-82bc-e58213caa38e_story.html)

Viki, G. T., & Abrams, D. (2002). But she was unfaithful: Benevolent sexism and reactions to rape victims who violate traditional gender role expectations. *Sex Roles, 47*, 289–293

West, K., Hotchin, V., & Wood, C. (2017). Imagined contact can be more effective for participants with stronger initial prejudices. *Journal of Applied Social Psychology, 47*(5), 282-292.

Wilson, T. D. (2004). *Strangers to ourselves*. Harvard University Press.