Supporting Information (Online Appendix) for "Backlash to Policy Decisions: How Citizens React to Immigrants' Rights to Demonstrate"

A) Re-Casting the Backlash Argument in Terms of Potential Outcomes

In the following we give a more formal description of our hypothesized causal mechanism. We define more precisely how we think about the policy effect on citizens' attitudes toward Muslim immigrants and lay open the assumptions needed to identify and estimate this effect.

Let $Y_i(t)$ denote the potential value of the outcome for unit i under the treatment condition $T_i = t$. In our context, this specifies a citizen's attitude toward Muslim immigrants, given a specific policy decision by the authorities (e.g. whether the authorities decide to permit or ban a public demonstration of Muslim immigrants), where $Y_i(t = 1)$ is how the citizen feel toward Muslims under a liberal policy and $Y_i(t = 0)$ how the exact same citizen would feel toward Muslims under a restrictive policy decision by the authorities. The causal policy effect is then captured by $\tau_i(t) \equiv Y_i(t = 1) - Y_i(t = 0)$, the difference in citizen *i*'s opinion under liberal and restrictive

regulation.[1] Of course, we can only ever observe one of *i*'s potential opinions at a given point in time, and since we are usually interested in the whole population, the actual quantity of interest is the *average treatment effect* (ATE): $\bar{\tau}(t) \equiv \mathrm{E}[Y_i(t=1) - Y_i(t=0)]$ across all citizens.

Since our theoretical argument stresses the importance of citizens' policy reaction, we now decompose this total policy effect in two parts: an indirect or mediated effect that runs via citizens' response – and thus captures our argument – and a direct effect that captures all possible remaining policy influences on citizens' attitudes (cf. Imai et al. 2011). Formally, we introduce $M_i(t)$ to denote the potential *mediator* of unit i under the treatment condition $T_i = t$, i.e. citizen *i*'s reaction to the authority's policy proposal. Thus, $M_i(t=1)$ is how the citizen would react to a liberal policy and $M_i(t=0)$ how the same citizen would react to a restrictive policy decision by the authorities. Then $Y_i(t, m)$ is the potential outcome if the treatment takes the value t and the mediator the value m. In our case, how a respondent feels toward a religious group under a particular policy and his or her particular reaction to this policy. We can now define the *indirect* or *causal mediation effect* as

$$\delta_i(t) \equiv Y_i\big(t, M_i(1)\big) - Y_i\big(t, M_i(0)\big), \tag{1}$$

for unit i and treatment status $t = 0,1$. We are thus interested in the difference in citizens' attitudes toward religious groups that would occur under liberal regulation and the attitude that would occur under the same condition but where citizens now react differently, namely as they would have reacted under restrictive regulation. Put differently, this is the effect a policy decision has on a citizen's attitude that is only due to the citizen's policy reaction. By fixing the policy and only changing the citizen's

[1] We have omitted possible covariates from the exposition for clarity.

response we isolate our hypothesized mechanism from all other possible mechanisms through which regulation may impact on attitudes (Imai et al. 2011). Again, we are interested in the *average causal mediation effect* (ACME) $\bar{\delta}(t)$, i.e. how the population thinks about Muslim immigrants compared to how the population would think about Muslim immigrants if we changed their policy reaction, while holding the actual policy constant.

While we argue that policy affects citizens' attitudes toward Muslim immigrants because of their support or opposition to elite policy decisions, there may exist alternative ways policies impact on citizens' attitudes. These alternatives are captured by the *natural direct effect*

$$\zeta_i(t) \equiv Y_i\big(1, M_i(t)\big) - Y_i\big(0, M_i(t)\big), \tag{2}$$

for unit i and treatment status $t = 0,1$. This is the change in a citizen's opinion on Muslim immigrants when changing the policy but holding his or her reaction constant. The *average natural direct effect* (ANDE) $\bar{\zeta}(t)$, therefore captures all policy effects that impact public opinion on Muslim immigration but which do not work through citizens' critical response.

Finally, we are also able to define a *controlled direct effect* as

$$\gamma_i(t) \equiv Y_i(1, M_i = m) - Y_i(0, M_i = m), \tag{3}$$

for unit i and for a fixed value for the mediator m. This effect captures the effect of changing the treatment while holding M_i constant and differs from the natural direct and indirect effect because it is defined in terms of *specific* values of the mediator and not in terms of the *potential* values of the mediator under different treatment conditions (see Pearl 2001, Imai et al. 2010a, Acharya et al. 2016). Whereas the indirect effect captures how the outcome changes with a treatment-induced change in the

mediator, the controlled direct effect captures the treatment-induced change in the outcome as a function of the value of the mediator: the first is the *mediating* effect of the mediator, the second is its *moderating* effect (Imai et al. 2010a). In terms of our research interest it captures how a specific degree of support or opposition to liberal policy affects the policy impact on attitudes toward Muslim immigrants. Again, we focus on the *average controlled direct effect* (ACDE) $\bar{\gamma}(t)$.

B) Identification

Leveraging the information obtained from this experiment, we are able to causally identify policy effects on citizens' attitudes toward religious groups under the so-called *sequential ignorability assumption* (Imai et al. 2010a, 2010b, 2011). This assumption states that, *first*, conditional on pre-treatment confounders, the treatment assignment is ignorable, i.e. statistically independent of both potential outcomes and potential mediators. *Second*, the assumption says that the mediator is ignorable, conditional on treatment status and pre-treatment confounders.

Our survey experiment clearly satisfies the first part of this assumption. Given both, the random assignment of the policy treatment and the large number of observations, we can safely assume that the two comparison groups are well balanced in terms of observed and unobserved pre-treatment variables. Table S1 supports this assumption with regard to key pre-treatment observables.

In contrast, citizens' policy reaction is not experimentally manipulated, but observed. Since we cannot control for all possible confounders between citizens' reaction and their feelings toward religious groups, as in all observational studies the second part of the sequential ignorability assumption is likely to be violated. Since the assumption

cannot be directly tested, we test the sensitivity of our results to this assumption in sensitivity analyses (Imai et al. 2010a, 2011). See supporting information E.

Tab S1. Balance in observables between treatment and control group

	Liberal Policy Condition	Restrictive Policy Condition	Difference	p-value
Respondent Characteristics				
Female	0.522	0.492	0.029	0.06
Age	50.998	50.785	0.213	0.67
Higher Education	0.453	0.467	-0.015	0.35
Left-Right-Ideology	4.029	3.947	0.082	0.08
Subjective Religiosity	2.786	2.768	0.018	0.84
Vignette Characteristics				
Christian	0.516	0.501	-0.015	0.34
Muslim	0.484	0.499	0.015	0.34
Non-practicing	0.335	0.335	-0.001	0.97
Devout	0.33	0.324	-0.005	0.71
Radical	0.335	0.341	0.006	0.68
Bulgarian Immigrant	0.339	0.327	-0.012	0.42
Nigerian Immigrant	0.331	0.332	0.001	0.96
Native Briton	0.329	0.341	0.011	0.45

Note: p-values from two-sided t-tests.

C) Description of Simple Experimental Results

In this section, we provide a brief description of the outcomes of our survey experiment. For each of the 36 treatment combinations (i.e. for 3 immigrant groups x 2 religions x 3 types of religiosity x 2 policy decisions), we look at the average policy reaction (Table S2) as well as the average feeling thermometer scores (Table S3). We structure the discussion along the two key comparisons between liberal and restrictive policy (permits vs. bans) and between Muslim and Christian groups.

Table S2: Support for the Policy Decision to Either Ban or Permit Public Demonstrations

		Non-practicing			Devout			Radical		
		Ban	Permit	Diff.	Ban	Permit	Diff.	Ban	Permit	Diff.
Native Britons	Christian	2.6	3.3	+.07*	2.4	3.5	+.11*	3.0	3.0	+/-.0
	Muslim	2.9	3.0	+.01	2.9	2.9	+/-.0	3.9	2.2	-1.7*
	Diff.	+.03	-.03		+.05*	-.06*		+.09*	-.08*	
Immigrants: Bulgaria	Christian	3.3	2.5	-.08*	3.0	2.8	-.02	3.5	2.4	-1.1*
	Muslim	3.4	2.5	-.09*	3.1	2.6	-.05*	3.8	2.3	-1.5*
	Diff.	+.01	+/-.0		+.01	-.02		+.03	-.01	
Immigrants: Nigeria	Christian	3.2	2.5	-.07*	3.0	3.1	+.01	3.5	2.4	-1.1*
	Muslim	3.0	2.6	-.04	3.2	2.5	-.07*	4.0	2.1	-1.9*
	Diff.	-.02	+.01		+.02	-.06*		+.05	-.03	

Note: Means on a 5-point scale reported. Higher values indicate higher support to the policy decision. * Bonferroni corrected $p < .05$.

Overall, respondents favor restrictive rather than liberal policy toward the right to demonstrate. Support for bans is significantly higher than for permissions in no less than 10 out of 18 total comparisons (we account for multiple comparisons by relying on the Bonferroni correction). In fact, we only find positive effects of liberal policy decisions[2] for native Britons and only if the group is described as secular or devout Christian (+0.7* and +0.11*, respectively), but not for native Muslims or radical groups. Perhaps unsurprisingly, preferences for allowing the demonstrations of native radical Muslims are significantly and markedly less frequent than for their banning (-1.7*). This particular group is likely viewed as highly problematic if not outright dangerous to British society.

The degree of religiosity seems to be more important than religious faith per se. Whether a group is described as Muslim or Christian does matter, but it is less relevant for citizens' policy reaction than might be expected. Differences between these two faith groups are only significant in 5 out of 18 comparisons, four of which refer to natives, where Christian demonstrations are generally met with more support than Muslim demonstrations. But this holds only for devout and to a stronger degree for radical groups, not for secular groups. When it comes to immigrants, citizens do not differentiate between Muslim and Christian groups at all (nor between Bulgarians or Nigerians for that matter). This is likely due to the fact that immigrants' right to demonstrate is generally met with low support in the first place. The only exception to this general pattern are devout Nigerians were respondents are again significantly less likely to support public demonstrations for Muslims (-.08*).

[2] These effects are simple *sample average treatment effects* (SATE), i.e. differences in means between the liberal and the restrictive policy conditions.

Turning to the feeling thermometer scores we find no noteworthy differences between the liberal and the restrictive policy decision conditions (Table 2). Not only are the sizes of the differences generally small, they also do not reach conventional levels of significance for any of the 18 comparisons. While this finding indicates that there is no direct effect of policy decisions on respondents' feelings toward religious groups, it leaves open the idea that any such effect depends on citizens' policy reaction (which we will address head-on in the following section). We also fail to find notable differences in citizens' feelings toward Muslims and Christians. Muslims are met with significantly cooler feelings in only 3 out of 18 comparisons and all of these instances refer to radical religious groups. Secular or devout Muslims are not significantly less liked than secular of devout Christians. Again, and perhaps unsurprisingly, the by far least liked groups are the radical religious.

Table S3: Feeling Thermometer Scores Toward Religious Group

		Non-practicing			Devout			Radical		
		Ban	Permit	Diff.	Ban	Permit	Diff.	Ban	Permit	Diff.
Native Britons	Christian	54	53	-1	53	50	-3	30	30	+/-0
	Muslim	51	51	+/-0	50	48	-2	20	18	-2
	Diff.	-3	-2		-3	-2		-10*	-12*	
Immigrants: Bulgaria	Christian	41	41	+/-0	49	47	-2	23	28	+5
	Muslim	43	40	-3	44	43	-1	20	19	-1
	Diff.	+2	-1		-5	-4		-3	-9	
Immigrants: Nigeria	Christian	40	37	-3	50	47	-3	27	24	-3
	Muslim	46	43	-3	41	37	-4	14	18	+4
	Diff.	+6	+6		-9	-10		-13*	-6	

Note: Means on a scale from 0 to 100 reported. Higher values indicate warmer feeling toward the group. * Bonferroni corrected p <.05.

In the analyses in the main text we collapse the answers to the "ban" and "permit" manipulation. This could be problematic if respondents attached vastly different meanings to a "ban" vs. a "permit". Yet if this were the case, we would expect to see the following in our data:

a) Reflecting the difference, there should also be a pronounced difference between responses "supporting a ban" and "opposing a permit". If there is no such difference, respondents are likely to attach symmetric meanings to these events. Table S4 compares the response distributions across the two experimental groups in our data set. As becomes apparent there are only tiny deviations in the responses and within the margin of sampling error. A chi-squared test for the null hypothesis of no differences across the two groups yields a p-value of 0.22. We interpret this none-difference across manipulations as suggesting that respondents don´t attach worrisome additional meanings to the "ban" of a demonstration.

b) If respondents indeed associate bans with the threat of violence from protestors as suggested by the reviewer, we would expect that they also attach different feelings to the demonstrating groups depending on whether the demonstration was banned or permitted. This information is readily available in Table S3 which displays the feeling thermometer scores across all groups and tests for differences between the "ban" and "permit" manipulations. In a total of 18 comparisons we find not a single significant difference in feelings toward "banned" and "permitted" demonstrating groups. Thus, respondents are unlikely to see them as more "dangerous" or "violent" just because of the manipulation.

Table S4: Distribution of responses to different experimental manipulations

"Ban"		"Permit"	
Strongly support the ban	0.21	Strongly oppose the permit	0.22
Support the ban	0.19	Oppose the permit	0.21
Neither support nor oppose the ban	0.30	Neither support nor oppose the permit	0.31
Oppose the ban	0.22	Support the permit	0.21
Strongly oppose the ban	0.08	Strongly support the permit	0.06

D) Model estimation

The statistical analysis of the experimental data and the estimation of the different causal effects (ACMEs, ANDEs, and ACDEs) proceed in the following steps. We first assess whether liberal or restrictive policy decisions by the authorities (T) have a causal effect on citizens' response and evaluation of a group's right to hold public rallies and demonstrations (M) in a simple regression equation

$$M_i = \alpha_1 + \beta_1 T_i + \xi_1{}' X_i + \epsilon_{1i}, \tag{4}$$

where X_i contains basic pre-treatment covariates (sex, age, education, political ideology and religiosity)[3] and ϵ_{1i} is the usual error term.

In a second step, we then test whether and how this reaction mediates the policy effect on citizens' general feeling toward that group (Y) in the equation

$$Y_i = \alpha_2 + \beta_2 T_i + \gamma M_i + \theta T_i M_i + \xi_2{}' X_i + \epsilon_{2i}, \tag{5}$$

which also includes a multiplicative interaction term between treatment and mediator, basic pre-treatment covariates, and an error term.

Using the estimates from these two equations, we then employ the algorithm proposed by Imai et al. (2010a, b) to calculate the ACMEs and ANDEs as defined above. This is done by first predicting the policy response under the two policy conditions (keeping all other variables at their empirical values) and then plugging these predicted values into the predictive formulas for the feeling thermometer scores, again setting the policy

[3] These covariates are coded as follows: sex (1- female, 0 – male), age in years, education (1 – university degree and higher, 0 – less than university degree), left-right-ideology („*In politics people sometimes talk of 'left' and 'right'. Where would you place yourself on this scale, where 1 means the left and 7 means the right?*"), and subjective religiosity („*Regardless of whether you belong to a particular religion, how religious would you say you are? 0- not at all religious to 10- very religious*").

conditions to its two values (and again leaving the remaining covariates at their empirical values). The propagation of inferential uncertainty from the first to the second equation and the resulting 95 percent confidence intervals for the ACMEs and ANDEs are obtained by running s = 1000 quasi-Bayesian simulations.[4]

We estimate the ACDEs using the sequential g-estimation algorithm proposed by Acharya et al. (2016). In the first stage, we start with the above regression equation (5). All model terms that involve the mediator (i.e. $\gamma M_i + \theta T_i M_i$) will be used as *demediation function*. This is then subtracted from the outcome of the second stage equation, which regresses the feeling thermometer scores on the policy treatment and the pretreatment covariates:

$$(Y_i - \hat{\gamma} M_i - \hat{\theta} T_i M_i) = \alpha_3 + \beta_3 T_i + \xi_3{}' X_i + \epsilon_{3i}, \qquad (6)$$

where the ACDE is now simply β_3. We re-center the demediation function to five different values k of the mediator, i.e. $\hat{\gamma}(M_i - k) + \theta T_i (M_i - k)$, and re-estimate the second stage equation each time to obtain ACDEs conditional on specific values of policy response. Again, to propagate the uncertainty from the first to the second stage and to construct 95 percent confidence intervals for the ACDE we run s = 1000 quasi-Bayesian simulations.

We deal with missing data in key covariates (respondent's education, religiosity, and left-right-ideology) by means of multiple imputation using chained equations (Little and Rubin 1987). We imputed five complete data sets and ran all models on each of these data sets. We present the combined results.

[4] These simulations are „quasi-Bayesian" in the sense that we are not specifying and sampling from a fully Bayesian model with prior distributions.

E) Model Results

Model 1 in table S5 presents the results of the first regression equation for the effect of permitting demonstrations on respondents' policy reaction. A simple comparison of liberal and restrictive authority decision reveals that a liberal decision gains significantly *lower* support from respondents (-.10*), although the substantive effect size is not particularly great.

Respondents differentiate between groups and are less likely to support public demonstrations of Muslims than of Christians (-.33**). However, they are even less likely to support public demonstrations of immigrants (-.42**). Radical religious groups are met with the least amount of support (-.45**). This does not seem to be an expression of anti-religious sentiment. Indeed, devout groups are more readily granted the right to hold public demonstrations, than the non-practicing (.12**) who are the reference category.

Respondents themselves also differ in their policy reaction. While both females and older respondents tend to be less supportive (-.19** and -.09**, respectively), higher levels of education are clearly related to more support (.34**). Ideologically right-leaning respondents are less likely to support a liberal policy decision toward public demonstrations (-.18**) and more religious respondents are more likely to support a demonstration (.02*). Adjusting for these pre-treatment covariates leaves the coefficients vignette characteristics unchanged.

Having demonstrated that policy decisions have a causal effect on citizens' policy reaction, we now seek to understand whether citizen reaction triggers a change in more general group evaluations. Model 2 presents the results of the second regression equation for respondents' feeling thermometer scores, which includes citizens' reaction and also interacts it with the policy decision. As before we find that

respondents have cooler feelings toward Muslims (-1.80**), immigrants (-3.07**), and especially radical religious groups (-18.71**). In general, women (3.42**), older (.83**) and more religious respondents (1.01**) have warmer feelings, and political right leaning respondents have cooler feelings (-.91**). But most importantly for our present purposes, we find a conditional effect of liberal policy decisions on feeling thermometer scores that depends on respondents' reaction (1.92**). The negative impact of liberal policy decisions becomes weaker and eventually turns positive as citizens' support increases.

Based on these two regression equations we can now derive our causal quantities of interest as defined above.

Table S5: Regression models of the causal effect of policy decisions (T) on citizen reaction (M) and general feelings toward religious groups (Y).

	M1: *Citizen Reaction* (M)	*M2:* *Feeling Thermometer* (Y)
Liberal Policy (T)	-.10*	-5.16**
	(.04)	(1.53)
		5.28**
Citizen Reaction (M)	--	(.81)
Liberal Policy (T) x Citizen Support (M)	--	1.92**
		(.52)
Vignette Characteristics		
Muslim	-.33**	-1.80**
	(.04)	(.64)
Devout	.12**	0.64
	(.05)	(.78)
Radical	-.45**	-18.71**
	(.04)	(.78)
Immigrant	-.42**	-3.07**
	(.04)	(.68)
Respondent Characteristics		
Female	-.19**	3.42**
	(.04)	(.63)
Age/10	-.09**	.83**
	(.01)	(.21)
Higher Education	.35**	-.66
	(.04)	(.68)
Left-Right-Ideology	-.18**	-.91**
	(.01)	(.24)
Subjective Religiosity	.02*	1.01**
	(.01)	(.11)
Intercept	4.48**	20.82**
	(.11)	(1.89)
N	3601	3601
R^2	.19	.41

Note: Unstandardized coefficients and standard errors in parentheses. Combined results from 5 multiply imputed data sets. * p<.05, ** p <.01

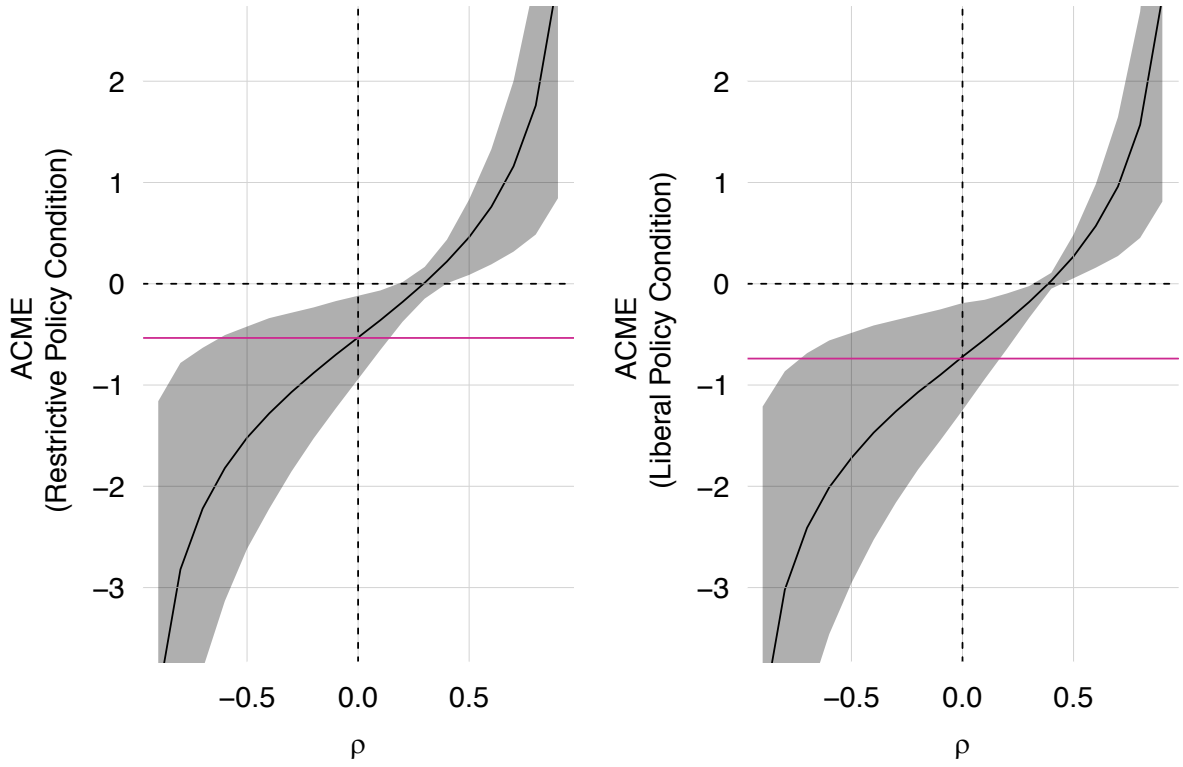
F) Sensitivity Analysis

Since the causal interpretation of policy effects rests on the sequential ignorability assumption and in particular the untestable assumption of no unobserved confounders of the relation between citizens' reaction (M) and their feelings toward religious groups (Y), we conducted a sensitivity analysis to assess the sensitivity of our results to this assumption (Imai et al. 2010a). One way to go about this sensitivity analysis is to understand how such an unobserved pre-treatment confounder would have to look like in order to change our inference. Following Imai et al. 2010a we rely on the sensitivity parameter ρ, the correlation between the errors of the regression equations for the mediator and the outcome variable, respectively. A non-zero correlation can be interpreted as a sign for the existence of an omitted variable and potential confounder of the relationship between mediator and outcome. One can now vary the values of ρ, calculate the corresponding ACMEs and check at which value the ACME becomes zero. This tells us what an unobserved confounder would need to look like in order to doubt our results.

Figure S3 presents this sensitivity analysis for the ACMEs under the control and the treatment condition by plotting them (along with their simulated 95 percent confidence intervals) against ρ. The solid horizontal line indicates the estimated ACME under the assumption of ρ=0. We find that under the restrictive policy condition the mediated effect remains robust unless ρ is greater than .3 and the effect under the liberal policy condition even until ρ is greater than .4. This indicates that our finding of a negative mediated effect of liberal policy on general feelings toward religious

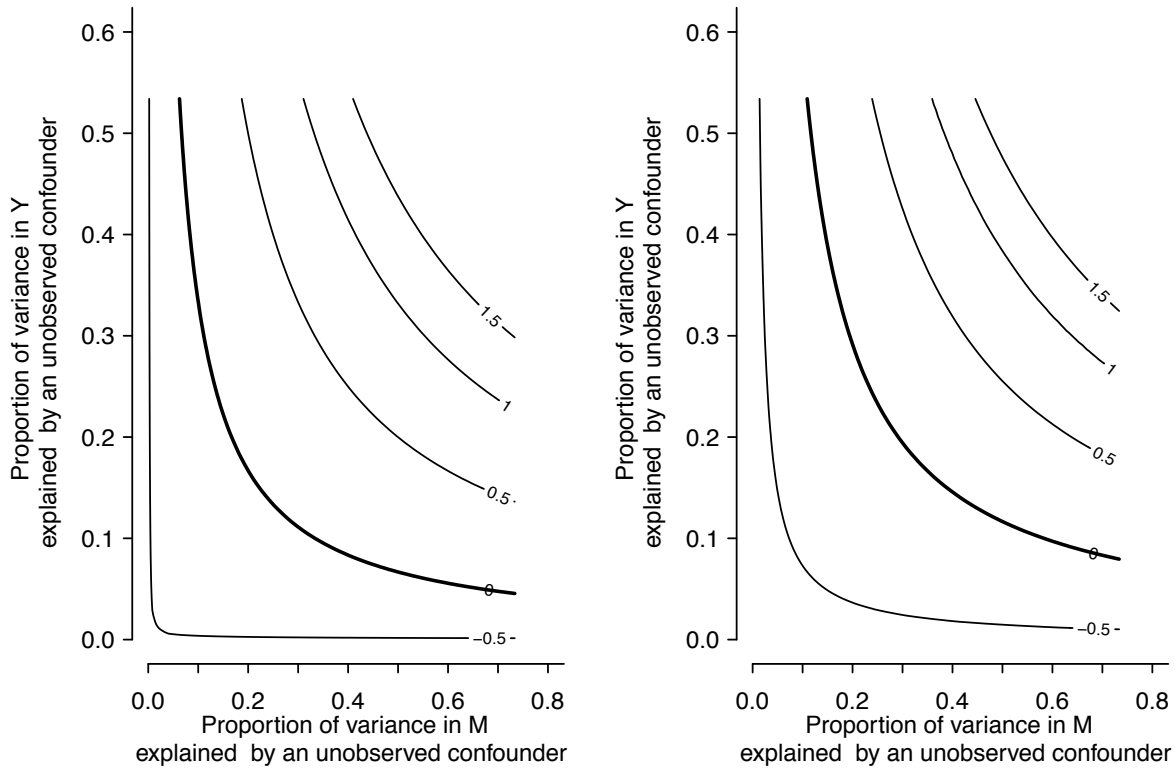
groups is relatively robust to a considerable deviation from the assumption of a zero-correlation between the errors of the mediator and outcome equations.[5]

Figure S3: Sensitivity Analysis of the sensitivity of the ACMEs to unobserved confounders of citizen reaction and feelings toward religious groups. The sensitivity parameter is the correlation rho between mediator and outcome errors.



[5]For related experimental studies on attitudes toward immigrants (Brader et al. 2008) or on civil liberties and tolerance (Nelson et al. 1997) similar ρ values of .43 and .48 have been reported (see Imai et al. 2011: 776).

Figure S4: Sensitivity Analysis of the sensitivity of the ACMEs to unobserved confounders of citizen reaction and feelings toward religious groups. The sensitivity parameter is the proportion of total variance explained in the mediator and outcome.



An alternative and more readily interpretable definition of this sensitivity parameter is in terms of the proportion of the variance of the mediator and the outcome that is explained by an unobserved confounder (Imai et al. 2010a). Figure S4 presents contour plots of how the ACMEs depend on the proportion of the variance of citizens' reaction (M) explained by an unobserved confounder (on the x-axis) and the proportion of the variance of feeling thermometer scores (Y) explained by an unobserved confounder (on the y-axis). Under the restrictive policy condition the ACME is guaranteed to be negative as long as the product of the R^2 of citizens' policy reaction and the R^2 of citizens' feelings that is due to an unobserved confounder does not exceed .04. Under the liberal policy condition the ACME remains negative if this product does not exceed

.08. Thus, as long as an unobserved confounder does not explain more than eight percent of the variation in citizens' policy reaction it could explain any proportion of the variance of the feeling thermometer scores (and vice versa) and still not change our inference. While an R^2 of .08 may not sound like much it is helpful to note that, in our data and taking citizen reaction as dependent variable, respondents' age only has a R^2 of .03, higher education of .04, and ideological self-placement only of .07.[6] Overall, we therefore conclude that the violation of the second part of the sequential ignorability assumption is unlikely to have major consequences for our main inference.

[6] Note that this sensitivity analysis assumes that the unobserved confounder affects policy reaction and feeling thermometer scores in the same direction. If we assumed that the unobserved confounder had different effects on the mediator and outcome, the ACMEs would always be negative and therefore extremely robust.

G) Further Consequences for the Social and Political Rights of Immigrants

In this section we show that citizens' reactions to policy decisions not only affect their general feelings towards religious groups but have further consequences regarding the social and political rights of those groups. Table S6 presents the model equations relating the policy treatment of authority decision (T) and the mediating citizen response (M) to citizens' opinions on welfare deservingness, the right to vote, and the right to hold public office, respectively.[7] While respondents do not discriminate amongst religious traditions and type of religiosity when it comes to welfare deservingness – only immigrants are viewed as less deserving – they are less willing to extend the right to vote and to hold public office to the non-practicing and radical. And while immigrants are less likely to be granted those democratic rights, respondents do not distinguish between Muslims and Christians. But we find the same significant interactions between authority permission and citizens' policy reaction, which we found for the feeling thermometer scores. And this holds for all three dependent variables. The effect of liberal policy on citizens' attitudes clearly depends on whether they support or oppose the authority's decision.

[7] The wording of these survey questions is as follows. Welfare deservingness: „*Assume that a single mother from this group with two children is unemployed. To what extent do you think she deserves help from the government? Very undeserving of help from the government (0) to Very deserving of help the government (10)*". Right to vote: „*To what extent do you agree that the members of this group should be allowed: to vote in national elections? Agree strongly (1) to Disagree strongly (5)*". Right to hold public office: „*To what extent do you agree that the members of this group should be allowed: to hold public office? Agree strongly (1) to Disagree strongly (5)*". We reversed the answer categories of the last two items so that higher scores indicate higher levels of agreement. Again we chose to model those responses using OLS regressions for ease of interpretation and more straight forward sensitivity analyses.

Table S6: Regressions models of the mediated causal effect of political regulation (T) on attitudes toward welfare deservingness, the right to vote, and the right to hold public office.

	Welfare Deservingness (Y)	*Right to Vote* (Y)	*Public Office* (Y)
	M1 (OLS)	M2 (OLS)	M3 (OLS)
Liberal Policy (T)	-.43*	-.24*	-.27**
	(.20)	(.10)	(.09)
Citizen Support (M)	.65**	.32**	.38**
	(.10)	(.05)	(.05)
Liberal Policy (T) x Support (M)	.19*	.10**	.09**
	(.07)	(.03)	(.03)
Muslim	-.08	-.05	-.04
	(.08)	(.04)	(.04)
Devout	.11	.18**	.15**
	(.10)	(.05)	(.05)
Radical	-.09	-.01	-.42**
	(.10)	(.05)	(.05)
Immigrant	-1.15**	-1.41**	-1.10**
	(.09)	(.04)	(.04)
Female	.13	-.14**	-.12**
	(.08)	(.04)	(.04)
Age/10	-.07**	-.05**	-.02**
	(.03)	(.01)	(.01)
Higher Education	.24**	.25**	.28**
	(.09)	(.04)	(.04)
Left-Right-Ideology	-.56**	-.15**	-.13**
	(.03)	(.02)	(.02)
Subjective Religiosity	.05**	.01	.01
	(.02)	(.01)	(.01)
Intercept	5.72**	3.85**	3.29**
	(.40)	(.20)	(.19)
N	3286	3286	3286
R^2	.36	.45	.47

Note: Unstandardized coefficients and standard errors in parentheses.

Figure S5 presents the average causal mediation effects (ACME) of political regulation on attitudes toward a) welfare deservingness of religious groups, b) their right to vote, and c) their right to hold public offices. As there are no significant ANDEs we can rule out any alternative causal mechanisms that do not run via citizens' reaction to regulation. The ACMEs clearly echo the results found for the feeling thermometer scores: liberal policy has a negative effect on citizens' attitudes toward religious groups, which is entirely mediated via their negative reaction to the permissive regulatory decision of the authorities. The ACME of liberal policy decision on attitudes of welfare deservingness is -0.08 [-0.14, -0.02]. The ACME of political regulation on citizens' willingness to extend the right to vote to religious groups is -0.04 [-0.06, -0.02]. Finally, the ACMEs on citizens' opinion on the groups' right to hold public offices is -0.04 [-0.08, -0.00]. As before, the effect sizes are quite small in substantive terms. Nonetheless, they are reasonably robust and suggest that liberal policy toward leads to a negative reaction among citizens, which in turn will make them less generous toward the groups benefitting from the policy. When authorities are permissive, citizens are more likely to deny religious groups welfare benefits and the fundamental political rights of active and passive democratic participation.

Figure S5: ACMEs and ANDEs of policy decisions on citizen attitudes toward A) welfare deservingness of religious groups, B) their right to vote, and C) right to hold public office. Based on the results in table S6. 95 percent quasi-Bayesian confidence intervals are based on 1000 simulations.

Figure S6: ACDEs of liberal policy on attitudes on citizen attitudes toward A) welfare deservingness of religious groups, B) their right to vote, and C) right to hold public office. Based on the results in table 4. 95 percent quasi-Bayesian confidence intervals are based on 1000 simulations.

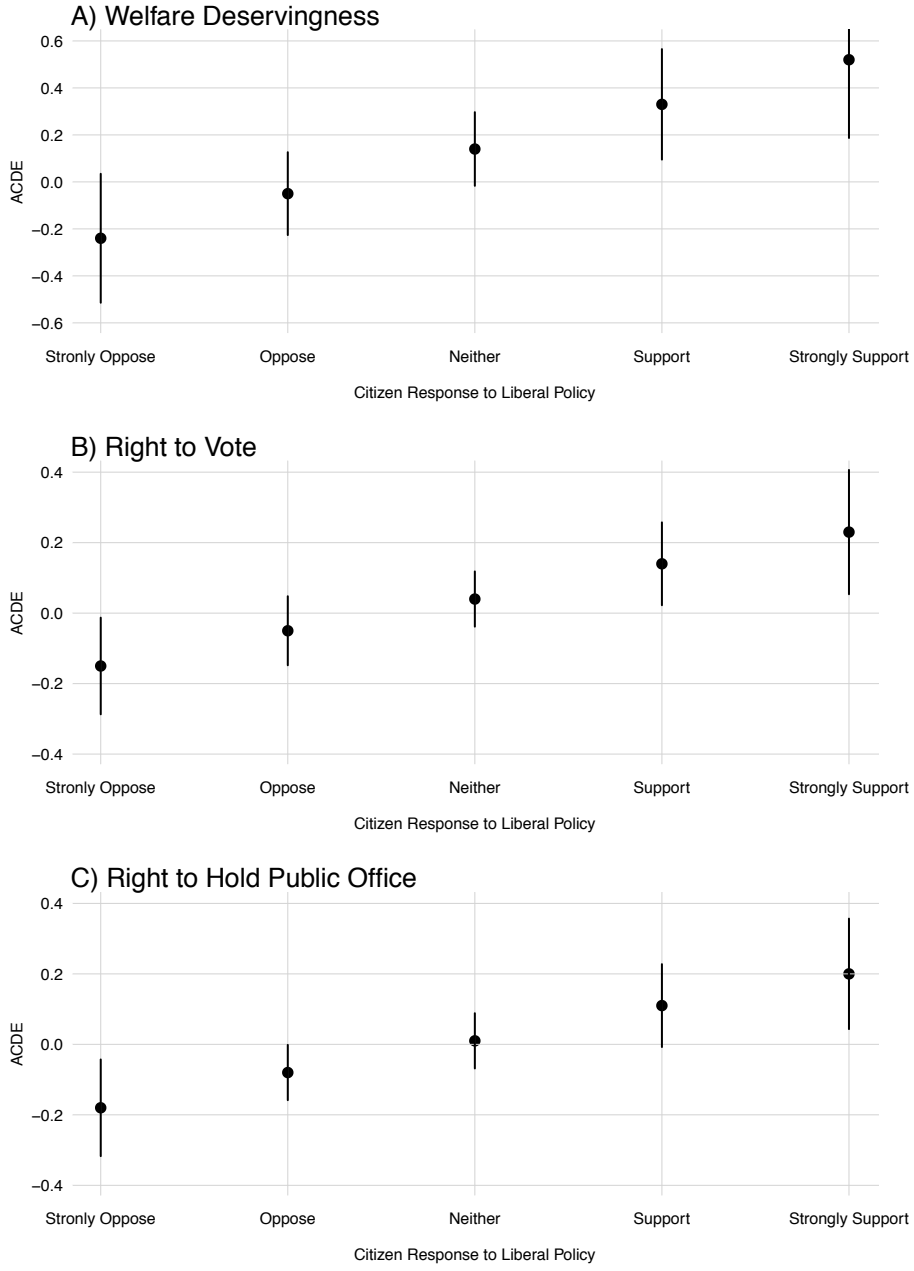


Figure S6 further illustrates the mechanism proposed by our backlash argument by plotting the ACDEs of liberal policy on citizens' attitudes toward the deservingness and democratic rights of religious groups, conditional on citizens' policy reaction. As in the case of the feeling thermometer scores, we find that liberal regulation has contradictory and polarizing effects, depending on whether citizens oppose or support this authority decision. Liberal regulation increases strong supporters' willingness to grant social rights (0.52 [0.19, 0.85]) as well as active (0.23 [0.05, 0.41]) and passive political rights (0.20 [0.04, 0.36]) to religious groups. At the same time those who strongly oppose liberal regulation are less likely to extend these rights when authorities adopt a permissive policy (welfare deservingness: -0.24 [-0.51, 0.03], right to vote: -0.15 [-0.29, -0.01], and hold public office: -0.18 [-0.32, -0.04]). This confirms the previously found pattern and generalizes it to a set of more specific social and political consequences for religious groups. The net result of this contradictory effect of liberal policy (as captured by the ACME) is negative: more citizens favor a restrictive approach and this preference is itself endogenous to permissive regulation by the authorities, resulting in overall negative attitudes toward religious groups.

Further results not presented here also generalize the previous finding that above all devout and secular religious groups suffer from opinion backlash.[8] We find that, under permissive regulation, citizens are less likely to grant devout Muslim groups welfare benefits, less likely to want them to vote, and to hold public office. The same trend is visible for non-practicing Muslims. Finally, opinions toward radical Muslims remain unaffected by liberal policy – citizens reject this religious group regardless of the specific policy approach.

[8] Full results for all religious groups, including Christians are available upon request. They largely echo the patterns found for the feeling thermometer scores.