Supporting Information (Online Appendix) for "Opinion Backlash to Policy Decisions: A Democratic Dilemma of Muslim Integration"

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

---

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

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 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	52.0	49.6	2.4	.10
Age	50.1	50.0	.1	.81
Higher Education	46.6	48.2	-1.6	.29
Left-Right-Ideology	4.0	3.9	.1	.10
Subjective Religiosity	2.9	2.9	.0	.96
Vignette Characteristics				
Christian	50.3	51.5	-1.2	.42
Muslim	49.7	48.5	1.2	.42
Non-practicing	32.6	33.0	-.4	.78
Devout	33.2	33.5	-.3	.87
Radical	34.2	33.6	.6	.65
Bulgarian Immigrant	33.2	34.2	-1.0	.47
Nigerian Immigrant	32.9	32.9	.0	.97
Native Briton	33.9	32.9	1.0	.49

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

C) 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)[2] 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 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.[3]

We estimate the ACDEs using the sequential g-estimation algorithm proposed by

---

[2] 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*").

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

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:

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

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.

D) Which Religious Groups Are Most Affected by Opinion Backlash?

In this section we demonstrate that the policy effect on citizens' feelings does not affect all groups the same way. Figure S2 presents the ACMEs of liberal policy on the feelings toward four religious groups cross-classified by religious tradition and immigrant status (these estimates are based on regression models with additional multi-way interactions, not shown here). In general, immigration status seems to matter more than religious tradition. While there are no significant mediated policy

effects on citizens' feelings toward UK natives, regardless of their religious identity, liberal regulation tends to produce cooler feelings toward immigrants. Whereas by strict adherence to conventional standards this negative effect is not statistically significant for Christian immigrants (treatment ACME: -0.59 [95% CI: -1.28, .10], control ACME: -0.33 [-0.72, 0.06], average ACME: -0.46 [-1.01, .09]), it clearly is for Muslim immigrants (treatment ACME: -1.05 [-2.03, -.07], control ACME: -0.77 [-1.50, -0.04], average ACME: -0.91 [-1.75, -0.07]). But note that the differences between the non-significant ACMEs for Christian immigrants and the significant ACMEs for Muslim immigrants are themselves not significant. This refines our previous result and suggests that citizens are especially critical of Muslim immigrants' public demonstrations and respond with greater dislike of this group when authorities decide to follow a permissive approach. As before, the substantive size of this policy effect is not very large but nonetheless visible.

Figure S2: Causal mediation analysis of the effect of regulation on citizens' feelings toward four religious groups cross-classified by religious tradition and immigrant status. ACMEs for treatment) and control condition as well as average ACME reported along with 95 percent quasi-Bayesian confidence intervals based on 1000 simulations.
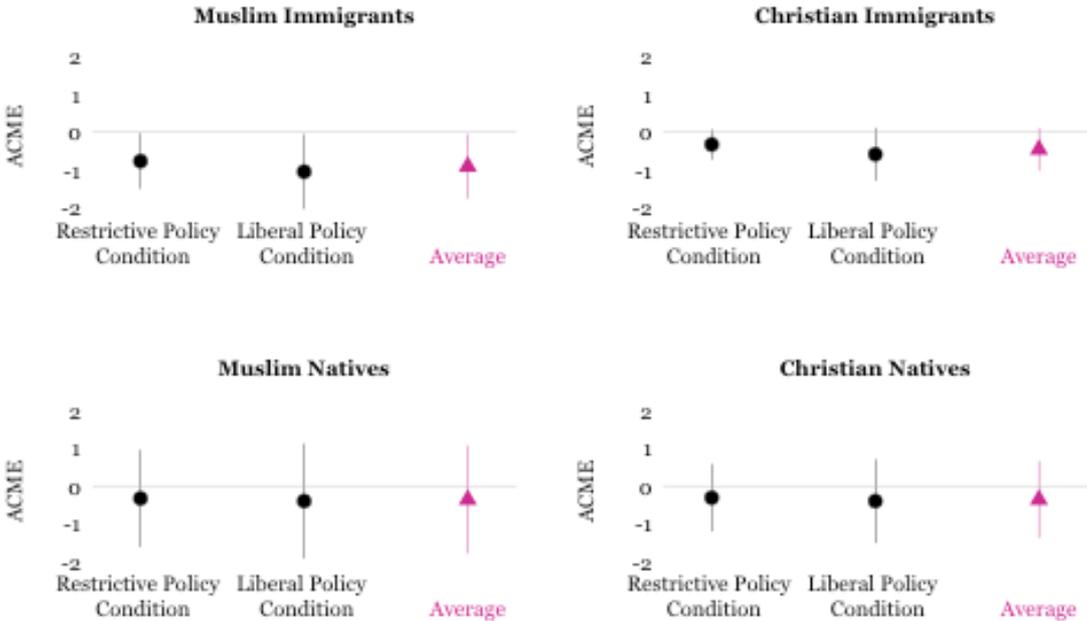
Next, we assess whether citizens reactions to liberal policy differ depending on the nature of Muslims' religiosity, i.e. whether the group is described as devout, radical or non-practicing. Figure S3 illustrates how permissive regulations affect citizens' feelings toward three Muslim groups classified by type of religiosity. The results suggest that devout and secular groups are more affected by citizens' backlash than radical religious groups. In fact, we found no significant policy effect on the feelings toward radical Muslims. Along with the low feeling thermometer scores reported in table 2 this hints at an unconditional rejection of religious radicalism, regardless of the policy approach adopted. However, permissive regulation leads to cooler feelings toward secular Muslims. The ACME for respondents in the liberal policy condition is -1.71 [-3.18, -0.24] and in the restrictive condition -1.59 [-2.98, -0.20] (average ACME: -1.65 [-3.08, -0.22]). Feelings toward devout Muslims are also reduced with the respective ACMEs of -2.01 [-3.48, -0.54] under the liberal policy condition and -0.54 [-1.52, 0.44] for the restrictive condition (average ACME -1.28 [-2.30, -0.26]).

Figure S3: Causal mediation analysis of the effect of regulation on citizens' feelings toward Muslim groups by type of religiosity. ACMEs for treatment and control condition as well as average ACME reported along with 95 percent quasi-Bayesian confidence intervals based on 1000 simulations.
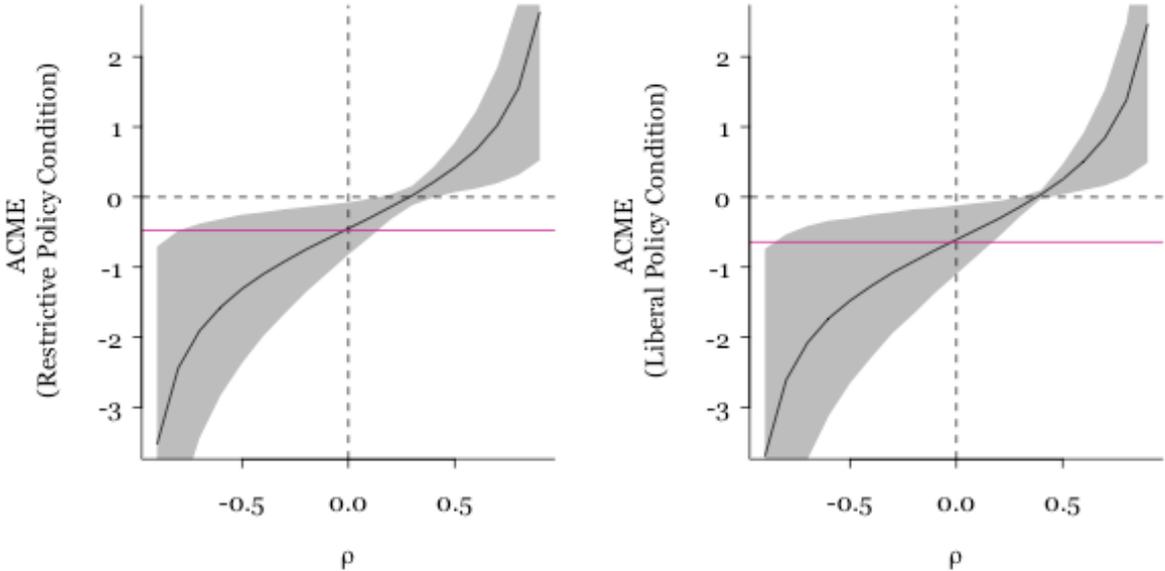
E) Sensitivity Analysis

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

10

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.[4]

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.

---

[4]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.
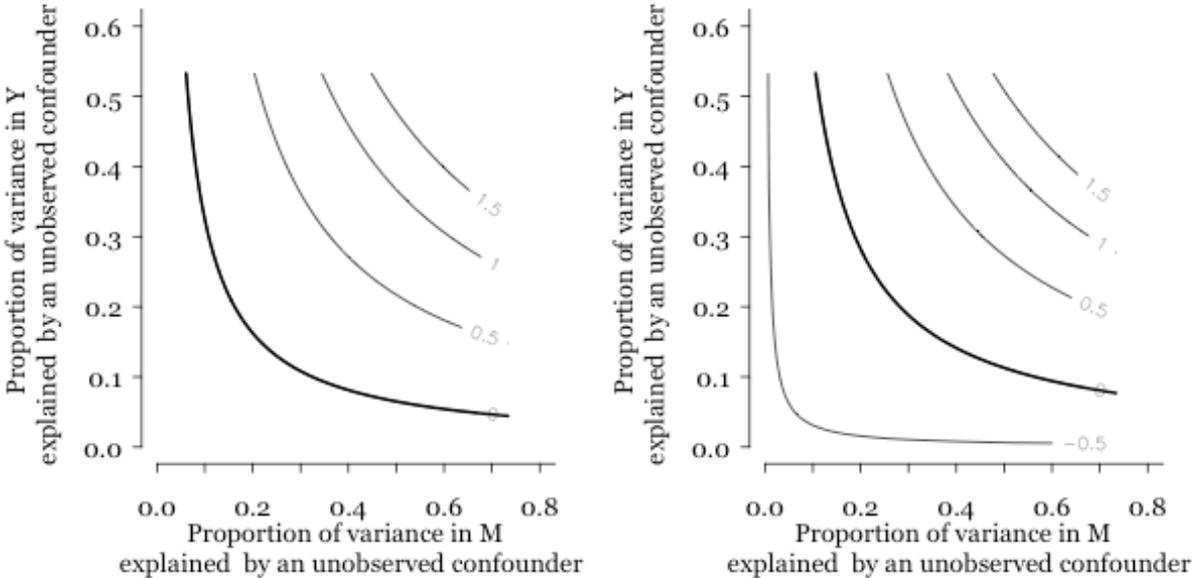
An alternative and more readily interpretable definition of this sensitivity parameter is in terms of the proportion of the variance of the mediator and the outcome that is explained by an unobserved confounder (Imai et al. 2010a). Figure S4 presents contour plots of how the ACMEs depend on the proportion 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 .05.[5] 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.

F) Further Consequences for the Social and Political Rights of Religious Groups

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 S2 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.[6] While respondents do not discriminate

---

[5] 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.

[6] 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

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.

---

chose to model those responses using OLS regressions for ease of interpretation and more straight forward sensitivity analyses.

Table S2: 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)	-.34 (.19)	-.19* (.09)	-.24** (.09)
Citizen Support (M)	.67** (.10)	.33** (.05)	.38** (.05)
Liberal Policy (T) x Support (M)	.15* (.07)	.08** (.03)	.08** (.03)
Muslim	-.07 (.08)	-.04 (.04)	-.03 (.04)
Devout	.12 (.10)	.20** (.05)	.16** (.05)
Radical	-.09 (.10)	-.02 (.05)	-.43** (.05)
Immigrant	-1.13** (.09)	-1.43** (.04)	-1.12** (.04)
Female	.08 (.08)	-.15** (.04)	-.14** (.04)
Age/10	-.09** (.03)	-.05** (.01)	-.03** (.01)
Higher Education	.24** (.09)	.23** (.04)	.25** (.04)
Left-Right-Ideology	-.56** (.03)	-.15** (.02)	-.13** (.02)
Subjective Religiosity	.06** (.01)	.01 (.01)	.01 (.01)
Intercept	5.76** (.38)	3.86** (.19)	3.33** (.18)
N	3565	3565	3565
R^2	.35	.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 -0.09 [-0.15, -0.03]. 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 -0.05 [-0.09, -0.01]. 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 S2. 95 percent quasi-Bayesian confidence intervals are based on 1000 simulations.

A) Welfare Deservingess

Causal Effect of Liberal Policy on Welfare Deservingness

B) Right to Vote

Causal Effect of Liberal Policy on Right to Vote

C) Right to Hold Public Office

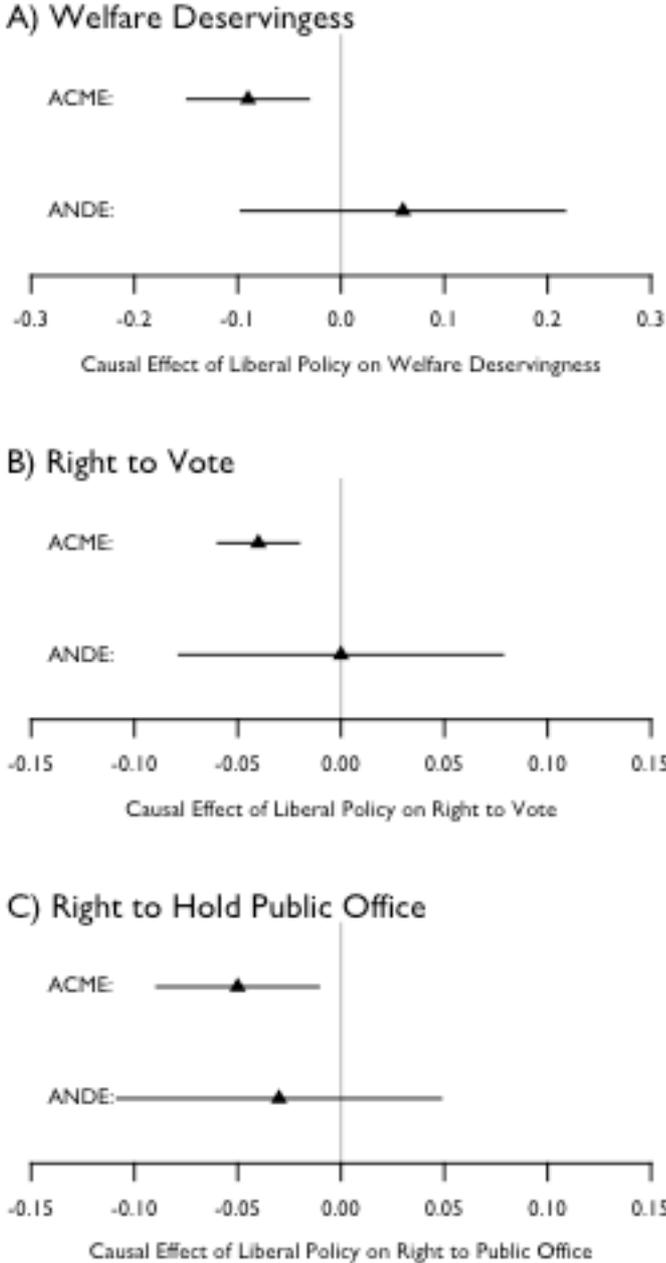Causal Effect of Liberal Policy on Right to Public Office

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.

A) Welfare Deservingness
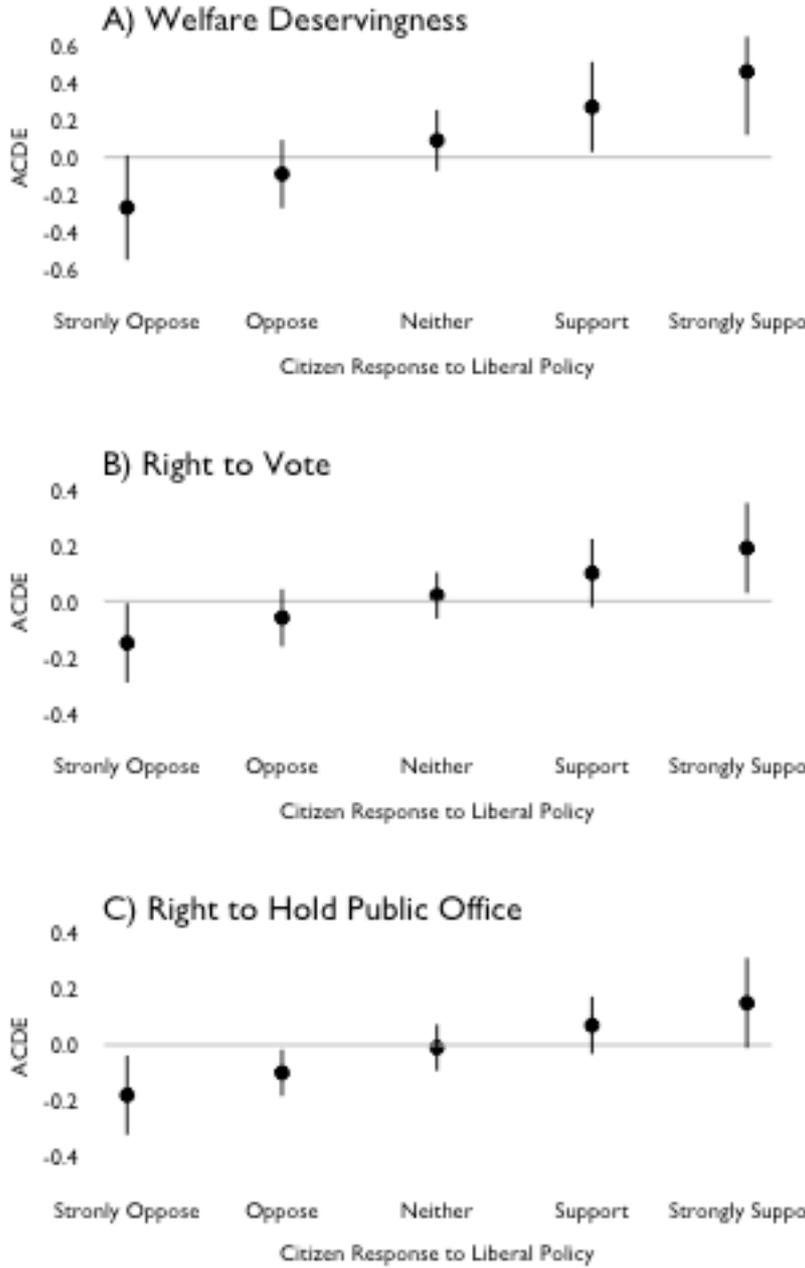

B) Right to Vote

C) Right to Hold Public Office

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 (0.46 [0.13, 0.79]) as well as active (0.19 [0.03, 0.35]) and passive political rights (0.15 [-0.01, 0.31]) 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: -.27 [-.54, .00], 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.[7] 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.

---

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