A. Elaborations on the estimation approach

In this section we elaborate on and clarify the estimation approach.

A.1. Estimation approach: zero-inflated beta regression

The spite score is a proportion between 0 and 1, and it represents how many points were taken away from the opponent relative to the maximally possible amount. A common approach for this kind of data is a beta regression.[40] A beta regression assumes that the data is distributed by a beta distribution—in contrast to the standard approach for non-proportion data which assumes a normal distribution. In particular, the problem with estimating proportions with a Gaussian distribution is that the support of the Gaussian distribution is the whole set of real numbers, while a proportion is a number between 0 and 1. Hence, the beta distribution (with support between 0 and 1) is the better estimation choice on theoretical grounds. In addition, we also discuss further estimation possibilities in C.5 and show that the beta distribution performs best in estimating the spite score. Hence, there are theoretical and practical reasons to explain why the beta regression is more appropriate than the normal distribution.

Another issue is that the spite score has an excessive amount of zeros, as can be seen in Figure 1. To deal with this, we use a zero-inflated beta model regression. The zero-inflated model estimates the decision to be spiteful in two parts: first, it estimates whether a participant decided to be spiteful or not (using a logistic regression); and second, it estimates conditionally on deciding to be spiteful, how spiteful participants decided to behave (using a beta regression). This allows us to use the data more efficiently and, more importantly, is less biased. The following example should clarify why a zero-inflated model gives a better picture than a model ignoring zero inflation. Imagine two very different groups, A and B. In group A, 80% of participants never behave spitefully, but 20% do, and these 20% behave 100% spitefully. In group B, 100% of participants behave 20% spitefully. If we would just use a simplistic approach and ignore the zero inflation, then we would find no significant difference between these two groups because both groups would show a spite score of .2 on average. However, this approach ignores essential information and does not reflect the situation adequately. In contrast, a zero-inflated model would show that group A is significantly more likely not to behave spitefully compared to group B and that conditional on the spiteful behavior participants of group A are significantly more likely to be fully spiteful.

[40] See for example Jönsson and Thor (2012); Laliberté et al. (2012); Rogers et al. (2012) for applications of the beta regression. For theoretical papers, see: Elgers et al. (1979); Grün et al. (2012); Schmid et al. (2013); Ospina and Ferrari (2012).

Hence, taking zero inflation into account gives us a clearer picture of the spite behavior.

Therefore, we use a zero-inflated beta regression to estimate our spite score. Nevertheless, we additionally include more ordinary estimation methods in Appendix D, where the main results can also be found together with the more common estimators, such as OLS and Tobit regressions. All results prevail if we use a standard linear regression or a Tobit regressions on the spite score. Thus, our results do not rely on the zero-inflated beta regression being used.

A.2. Estimation

We stipulate that our conditional spite score (y) follows a beta distribution with mean μ and precision σ, i.e. $y \sim \mathcal{B}(\mu, \sigma)$. Here, $\mathbb{E}(y) = \mu$ and $\mathbb{V}ar(y) = \mu(1+\mu)/(1+\phi)$. The beta density function is described by:

$$f(y; \mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1}$$

with $0 < \mu < 1$, $0 < y < 1$[41], and $\phi > 0$.

Let $y_1, ..., y_n$ be our observations of the conditional spite score, where every $y_i \sim \mathcal{B}(\mu, \sigma), 1 = 1, ..., n$.

Hence, the econometric model we use is described by:

$$\begin{aligned} g(\mu_i) &= \sum_{j=1}^{k} x_{ij}\beta_j \\ g(v_i) &= \sum_{j=1}^{k} x_{ij}\alpha_j \\ v_i^* &= 1 \text{ if } v_i > 0 \\ \mu_i^* &= \mu_i \cdot v_i^* \end{aligned} \tag{2}$$

where $\beta = (\beta_1, ..., \beta_k)^T$ and $\alpha = (\alpha_1, ..., \alpha_k)^T$ are vectors with the unknown regression parameters ($\beta, \alpha \in \mathbb{R}^k$) and $x_{i1}, ..., x_{ik}$ are the observations. $g(\cdot)\colon (0,1) \to \mathbb{R}$ is the link function. We will use the logit link: $g(\mu) = log(\mu/(1-\mu))$, as this is easier to interpret. We will assume $x_{i1} = 1$ to be the intercept.

[41] Note: Because y is stipulated to be strictly smaller than 1, we change the spite score of 1 to $1 - 10^{-5}$, as suggested in Smithson and Verkuilen (2006). Furthermore, because the conditional spite score is per definition bigger than 0, we do not need to adjust y at 0.

Our main model, which we denote by Ψ, estimates the Equation 2 with $x_{i1} = 1$ and $x_{i2} =$ Outgroup, where Outgroup is a dummy variable with value one if the opponent is an outgroup-member and zero if the opponent is an ingroup-member.[42] This model aims to directly answer the question of whether or not partisanship leads to dysfunctional behavior.

B. Participants and Demographics

B.1. Sample selection

We see that almost twice as many Clinton voters finished our experiment as Trump voters. However, the proportion of Clinton voters finishing our study is almost identical to the proportion of Clinton voters starting our study (62% vs. 58%). Looking at drop-out behavior, we find no relevant interaction effects between whom participants voted for and observables on the probability of not finishing the study. Thus, it does not look like we have a selection bias during the study. Yet, the question is whether we have a selective group of participants starting our study. The demographics reported in studies focusing on the general demographics of Mturkers (Robinson et al., 2019; Huff and Tingley, 2015; Difallah et al., 2018) are rather similar to ours. For example, Difallah et al. (2018) show that about 55% of US Mturkers are female, which is almost identical to the proportion of females in our study. Further Robinson et al. (2019), and Huff and Tingley (2015) show that there are substantially more democrats among Mturkers than republicans. Moreover, the recent study by Hyatt et al. (2018), which also elicits the vote in the 2016 election, estimates that the proportion of Clinton to Trump voters is 60% among Mturkers, which is very close to the 58% in our study. These insights indicate that self-selection into our study is rather unlikely.

B.2. Demographics

As with most experimental studies, our sample does not perfectly represent the American population.[43]

The age of our participants ranged from 18 to 88 years, with most participants in the age-group between 30 and 44 (49%) and 22%, 25% of participants in the

[42] Several additional robustness checks, such as including sociodemographic controls or taking information about crime, poverty, and religion of the county of the participants into account, can be found in Appendix C.4. The results prevail in all models.

[43] For comparison estimates see the census aggregates: https://www.census.gov/quickfacts/fact/table/US/PST045216 and https://www.census.gov/content/dam/Census/library/publications/2016/demo/p20-578.pdf.

age-group between 18 and 29, 45 and 64, respectively (Median = 36). Hence, our sample is younger than the average American with a median age of 37.9 and with 15% of the population older than 65 years (compared to 4% in our sample).
In addition, 53% of our participants were female compared to 50.8% females in the US population.

Concerning the ethnicity in our sample: 81% of participants are White compared to 61.3% Whites in the US population.

Moreover, our participants indicated to have a higher educational attainment than the typical American. 72% of participants implied to have at least a Bachelor's degree as the highest qualification compared to roughly 33% in the United States as a whole.

Hence, our sample is younger, more female, more white, and better educated than the average American.

In addition, looking at the location of the participants (see Figure 4), we find that the participants mainly come from populated and urban areas. This can also explain the discrepancy in the distribution of Trump and Clinton voters in our study (38% vs. 62%) compared to the distribution in the general election (46% vs. 48%).

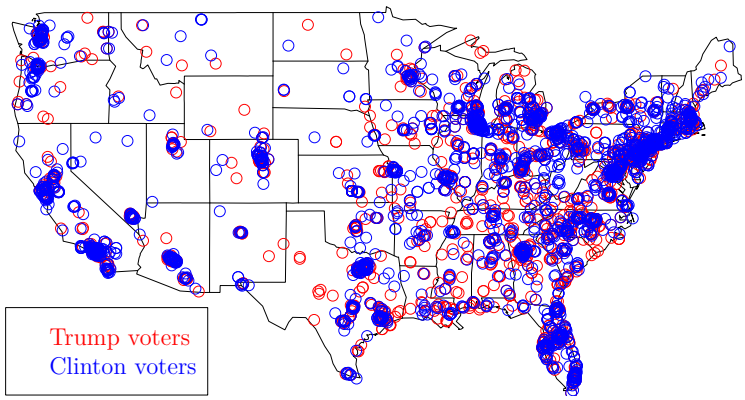


Figure 4: Participants' location by vote.
The figure depicts the location of our participants. Blue circles represent Clinton voters while red circles depict Trump voters.

B.3. Demographic voting patterns

Nevertheless, the participants in our study exhibit similar demographic voting patterns as reported in exit-pollings:[44] Trump voters in our sample are significantly less educated and are on average older than Clinton voters. Significantly more men voted for Donald Trump and significantly more white people voted for Trump than for Clinton. Table 4 shows the demographic differences between Clinton and Trump voters in our sample.

Test	Clinton voters	Trump voters	T	Df	p	95% CI	Sign.
Female	0.54	0.52	1.44	3533.00	0.15	[-0.01 , 0.06]	
Age	38.43	40.92	-6.00	3533.00	0.00	[-3.3 , -1.68]	***
Race=White	0.78	0.87	-7.00	3533.00	0.00	[-0.12 , -0.07]	***
College-Ed or Higher	0.78	0.69	6.25	3533.00	0.00	[0.06 , 0.12]	***
Income > $70k	0.45	0.48	-1.75	3533.00	0.08	[-0.06 , 0]	.
Notes:						·p<0.10;*p<0.05;**p<0.01;***p<0.001;	

Table 4: Demographics of participants.
The table depicts summary statistics on Clinton and Trump voters in our sample, and compare whether demographic characteristics differ between the two voter types. Female is a dummy with value one if the voting participant is a female and zero otherwise. Age is a continuous variable denoting the age of the voting participant. Race=White is a dummy with value one if the voting participant indicated to be white and zero otherwise. Similarly, College-Ed or Higher, Income > $70k is a dummy with value one if the voting participant has at least a college degree, annual income of more than $70k, respectively, and zero otherwise. We use two-sample t-tests to compare characteristics.

We see that Trump voters in our sample reflect Trump voters in the general election quite well. The same is true for Clinton voters in our sample compared to Clinton voters in the general election.

An analysis of the voter data just after the election revealed that "in the 2016 election, a wide gap in presidential preferences emerged between those with and without a college degree. College graduates backed Clinton by a 9-point margin (52%-43%), while those without a college degree backed Trump 52%-44%"(Pew Research Center, 2016). In our data, college graduates voted for Clinton 50% of the time while they voted for Trump 43% of the time. People without a college degree in our data voted for Clinton 41% and for Trump 57% of the time compared. Moreover, the analysis shows that "older voters (ages 65 and older) preferred Trump over Clinton 53%-45%."(Pew Research Center, 2016). In our data, the numbers are 55%-42%. In addition, women supported Clinton over Trump by 54% to 42% (Pew Research Center, 2016). In our data, the margin is 49%-45%. Further, young adults (18-25) preferred Clinton over Trump by a wide 55%-37%

[44] See Alcantara et al. (2016) or Kirk and Patrick (2016).

margin (Pew Research Center, 2016). In our data, the margin is 52%-40%. The analysis by the Pew Research Center (2016) also shows that "Trump won whites with a college degree 49% to 45%" and he won whites without a college degree 67% to 28%. In our data, Trump won whites with a college degree 46% to 48%, and he won whites without a college degree 61% to 39%.

Thus, our selected sample shows a striking similarity to the general populations' patterns and reflects the attitudes of general Clinton and Trump voters rather reliably.[45]

C. Robustness checks

C.1. Coherence

To ensure that participants are not randomly choosing a candidate and are really paying attention, we asked participants also for their preferred political party, and we included several attention checks.

C.1.1. Consistency

More specifically, we asked participants in the general demographics part "With which party do you normally identify yourself most with?" and later in the study, we asked "Which political party do you usually feel closest to".

99.4% of those participants who indicated to identify most with Democrats also felt closest to Democrats, and similarly, 99.1% of those participants who indicated to identify most with Republicans also felt closest to Republicans.

Furthermore, 91.9% of those participants who indicated to have voted for Hillary Clinton also felt closest to Democrats, and similarly, 84.5% of those participants who indicated to have voted for Donald Trump also felt closest to the Republicans.

Additionally, 77.7% of those participants who indicated to have voted for Hillary Clinton usually are identifying themselves with Democrats (6.2% usually identify themselves with Independents), and similarly 69.2% of those participants who indicated to have voted for Donald Trump usually identify themselves with Republicans (24.6% usually identify themselves with Independents).

[45] To deal with potential selection effects, we reestimate all regressions by reweighting our sample to make it more representative in Appendix D.2. All results prevail.

C.1.2. Attention Checks

In some of the questionnaires, we included additional attention checks by asking questions, for example, "Click on agree" or "This is another control. We ask you to select the second option." We included four of those attention checks (without having any impact on the participants). Only 2% of all participants failed one or more of those attention checks (some of the participants, however, reported to have misunderstood the meaning of "second option" as this might have been ambiguous in regard to the reference point).

Overall, participants seem to be attentive and consistent in their political attitudes in our study.

C.2. Consistency of the spite measure

In this subsection we provide some evidence that spite behavior is not driven by noise, and that it is not influenced by the lack of altruistic options.

C.2.1. Noise

In this section, we ask whether the spite behavior is driven by noise. One way to show that noise is very unlikely the main driver of the spite behavior in our experiment is by looking directly at individual choices in each of the three distribution decisions. If noise would be a relevant driver we would expect a sizable portion of choices to be somewhere in the middle. This is particularly true for the first and third distribution decision. For the second distribution, things are different as spiteful behavior is costly, and thus options in the middle are reasonable for a spiteful participant. However, if we were to assume that only spiteful preferences are driving behavior, we would expect participants to either choose the first option (no spite) or the last option (full spite) in the first and third distribution and would not expect participants to choose any other option. Obviously, there might be other reasons than noise driving participants to choose options in the middle, but for simplification, we just pretend as if all choices in the middle are due to noise. We see that for the very first distribution decision 93% of participants choose either no or full spite. We also see that for the third distribution decision 93% of participants choose either no or full spite. Thus, even if we conservatively assume that all those decisions in the middle are due to noise we see that noise, is basically negligible.

Further, we see that a vast majority of participants *consistently* do not choose options in the middle. Specifically, 93% of participants choose no option in the

middle for the first *as well as* for the third distribution decision. This indicates that a vast majority of decisions are not driven by noise.

Even if we were to drop those participants who choose an option in the middle in the first/third part, we still obtain almost identical results with the exception that the spite-score reduces slightly.

C.2.2. Lack of altruistic options

In this section, we ask whether adding altruistic options to the spite measure changes spite behavior. One potential issue with the current spite measure is that participants might be biased towards more spiteful behavior in our spite measure as all but the first choice are indicative of spiteful behavior. Thus, to test whether the lack of altruistic options in the spite measure is biasing behavior we conducted an extension experiment. The extension experiment consisted solely of the spite-task to measure whether adding altruistic options to the spite measure changes behavior. We had two main treatments in the extension experiment: a baseline treatment where the spite-task was identical to the one described in the main part of the paper with the exception that the third distribution decision was reversed and an extension-treatment where we added eight additional altruistic options. The spite-task with altruistic options is depicted in table 5.[46]

You receive	50	52	55	58	60	62	65	68	70	70	70	70	70	70	70	70	70
	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Other receives	115	113	111	109	108	106	104	102	100	98	96	94	92	91	89	87	85
You receive	50	52	55	58	60	62	65	68	70	68	65	62	60	58	55	52	50
	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Other receives	130	126	122	119	115	111	108	104	100	96	92	89	85	81	78	74	70
You receive	100	100	100	100	100	100	100	100	100	96	92	89	85	81	78	74	70
	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Other receives	85	87	89	91	92	94	96	98	100	102	104	106	108	109	111	113	115

Table 5: Spite measure with altruistic options.
The table depicts the nine allocation choices in each of the three decisions of participants in our own spite measure with eight additional altruistic options each (denoted by gray cells). For each choice, the upper row denotes the payoff in experimental currency units for the deciding participants, while the bottom rows each denote the payoff for the other player. Gray cells denote the additionally included altruistic options.

[46] Note that the altruistic options are costly on purpose. The main reason for this choice is that making the altruistic options costless would basically transform the altruistic options into spiteful options. For example, if the altruistic options of the first decision would be made costless, the decision-maker would always receive 70 points for any of the "altruistic" options. But then, the only non-spiteful option (and therefore also the only altruistic options) would be the 70-115 options, as any other choice would destroy the income of the opponent. Thus, to be able actually to offer altruistic options, they had to be made costly. Note further that the spiteful options in the second decision are also costly making the two option sets comparable.

The experiment was conducted in march of 2021. 271 participants finished this extension experiment.[47] We find that adding altruistic options to the spite measure slightly increases the spite score by 0.03, which, however is not significantly different from zero ($t(269)$= -1.1, $p \geq 0.05$). We also do not find a significant effect of adding altruistic options to the spite measure neither for Clinton voters ($t(149)$= -0.9, $p \geq 0.05$) nor for Trump voters ($t(102)$= -0.9, $p \geq 0.05$). Using more sophisticated estimations leads to the same results. Specifically, we estimate no significant effect of adding altruistic options to the spite measure using linear regressions, Tobit regressions as well as zero-inflated beta-regressions.

C.3. Feeling of warmth

In this section, we examine the properties of the feeling of warmth measure compared to results from the literature.

Notably, we asked participants for their feeling of warmth towards Trump and Clinton voters and additionally towards Republicans and Democrats. A recent study by the Pew Research Center also asked Republicans and Democrats for their feeling of warmth towards, among others, Republicans and Democrats. They also found a very similar gap. What is, however, striking is how close the estimations from our study are to the estimates of the study by the Pew Research Center (2017b).

Table 6 shows the estimates in our online study compared to the representative study by the Pew Research Center (2017b).

Even though our results are very similar to the results by the study of the Pew Research Center (2017b), our participants indicated to have, on average, lower feelings of warmth than participants in Pew Research Center (2017b). This might be partly explained by the fact that our experiment was conducted shortly before or after the election, and at that point, people were somewhat fed up with politics. Another explanation might be that we used a 10-point scale, and Pew Research Center (2017b) used a 100-point scale. Of course, it is also possible that our participants are particularly negative. However, the key observation from this

[47] In this extension experiment we manipulated two aspects: 1) we added eight additional altruistic options to the spite measure and 2) we randomly referred to the co-player either as the "opponent" or the "co-player". We find no significant effect in spite behavior of referring to the co-player either as the "opponent" or the "co-player" ($t(269)$= -1.2, $p \geq 0.05$). This null effect is found for Clinton ($t(149)$= 0.3, $p \geq 0.05$) and Trump voters ($t(102)$= -1.6, $p \geq 0.05$) alike. Also more sophisticated estimations find no significant change in behavior nor do we find any significant interaction effects by vote. Thus, we pool the data to have more power to test whether additional altruistic options change behavior.

comparison is that reported feelings of warmth in our online study are very similar to the representative study by the Pew Research Center (2017b).

	Attidues towards	Our study	Pew study
Among Democrats	Democracts	6.57 ± 0.1	7.10
	Republicans	1.5 ± 0.07	2.40
Among Republicans	Republicans	6.46 ± 0.14	7.00
	Democracts	2.36 ± 0.12	2.30

Table 6: Comparing feelings of warmth towards Democrats and Republicans in our study to the study by Pew Research Center (2017b).

C.4. Estimations with Controls

To ensure robustness of our results, we extend in this section the estimation of section 3.2 to further controls.

At first, we account for the responses in the spite questionnaire, the attitudes on the social distance measure, and the feeling thermometer.

Additionally, we control for socio demographics and relevant information of the participants' county/state. We control for socio demographics by: gender $\in$ { Male; Female }, Education $\in$ {College-ed. or higher; below College }, Age $\in \{18, \ldots, 88\}$, and reported Income $\in \{< 70k, > 70k\}$.

We also control for information about poverty, crime, and religion of the participants' county. The result can be found in Table 7.

The crime data was obtained from the "Uniform Crime Reporting Program Data: County-Level Detailed Arrest and Offense Data, 2014" reported by the United States Department of Justice, Federal Bureau of Investigation[48]. To control for crime in the county of the participating participant, we include the relative crime (amount of reported crimes relative to the population of the county) and the relative violent crime (amount of reported violent crimes relative to the population of the county).

The data on poverty on the county level was obtained from the United States Department of Agriculture[49] and we also obtained data on the state level from

[48]The data can be found here: http://doi.org/10.3886/ICPSR36399.v2 (Federal Bureau of Investigation, United States Department of Justice, 2016)

[49]The data can be found here: https://www.ers.usda.gov/data-products/county-level-data-sets/county-level-data-sets-download-data/ (United States Department of

the United States Census Bureau[50]. To control for poverty on the state level, we included the poverty as reported in (United States Department of Agriculture, 2017). We also controlled for the percentage of people living below the poverty level on the state level from the Census Data.

The data on how religious the county was, was obtained from the Pew Research Center 2014 U.S. Religious Landscape Study.[51] We controlled in two ways for how religious the county of the participant was. We included the percentage of participants answering yes on the question, whether they believe in god, and we secondly included the percentage of participants answering "very important" on the question: "How important is religion in your life".

Table 7 replicates Table 2 from section 3.2 while accounting for the above mentioned controls.

The resulting estimates show that the results are essentially robust to controls.

C.5. Model comparisons conditional spite

In this section, we compare further potential models to estimate the spite score against the beta regression. As the spite score has an excessive amount of zeros, it seems plausible to assume a two-step process while the first step is the decision to be spiteful or not. This decision is a binary decision, and hence, a logistic regression was used for this step. However, for the conditional spite score, we argued for a beta regression as this is the common method for this kind of data. To further support our claim, we compare modeling the spite score conditional on behaving spitefully with a beta regression compared to other regressions. It can be seen from Table 8 that the beta regression outperforms the other regressions (standard Gaussian-, Gamma-, Weibul-, log-normal-, exponential and exponential Gaussian regression) on the conditional spite score.[52]

Agriculture, 2017). The poverty estimates reported in this data are model estimates from the U.S. Census Bureau's Small Area Income and Poverty Estimate.

[50]The data can be found here: https://www.census.gov/data/tables/2016/demo/income-poverty/glassman-acs.html Glassman and United States Census Bureau (2016).

[51]The data can be found here: http://www.pewforum.org/datasets/pew-research-center-2014-u-s-religious-landscape-study/ (Pew Research Center, 2014).

[52]It is also noteworthy that the main results of this paper do not change, if one would use a different kind of model.

	Spite? Full sample	Score Full sample	Spite? Trump voters	Score Trump voters	Spite? Clinton voters	Score Clinton voters	Spite? Full sample	Score Full sample	Spite? Trump voters	Score Trump voters	Spite? Clinton voters	Score Clinton voters	Spite? Full sample	Score Full sample	Spite? Trump voters	Score Trump voters	Spite? Clinton voters	Score Clinton voters	Spite? Full sample	Score Full sample	Spite? Trump voters	Score Trump voters	Spite? Clinton voters	Score Clinton voters
Constant	−1.52***	0.29	−0.65·	−0.22	−1.73***	0.50*	−1.02***	0.13	−0.20	0.02	−1.15***	0.11	−2.21	0.81	−2.73	2.22	−1.27	−0.24	−4.37**	−0.07	−4.83·	1.34	−3.71·	−1.36
	(0.22)	(0.18)	(0.34)	(0.25)	(0.28)	(0.25)	(0.14)	(0.13)	(0.22)	(0.17)	(0.18)	(0.16)	(1.57)	(1.33)	(2.64)	(2.06)	(1.97)	(1.74)	(1.65)	(1.34)	(2.78)	(2.06)	(2.07)	(1.76)
Outgroup	0.49**	−0.08	−0.28	−0.01	0.67***	−0.23	0.50***	0.02	−0.13	−0.18*	0.50***	0.03	0.52***	0.01	−0.11	−0.15	0.52***	0.01	0.44**	−0.12	−0.32·	0.002	0.61**	−0.22
	(0.16)	(0.14)	(0.19)	(0.15)	(0.19)	(0.17)	(0.09)	(0.08)	(0.11)	(0.09)	(0.09)	(0.09)	(0.10)	(0.09)	(0.12)	(0.09)	(0.10)	(0.09)	(0.16)	(0.14)	(0.20)	(0.15)	(0.20)	(0.17)
	[11.35]	[-1.85]	[-6.55]	[-0.35]	[15.32]	[-5.68]	[12.10]	[0.56]	[-3.23]	[-4.49]	[11.90]	[0.67]	[12.54]	[0.25]	[-2.73]	[-3.59]	[12.34]	[0.23]	[9.86]	[-2.95]	[-7.24]	[0.05]	[13.56]	[-5.31]
Trump voter	0.59***	−0.17·					0.65***	−0.08					0.62***	−0.14					0.58***	−0.15				
	(0.12)	(0.10)					(0.11)	(0.10)					(0.11)	(0.10)					(0.12)	(0.10)				
	[13.72]	[-4.04]					[15.71]	[-1.91]					[14.88]	[-3.47]					[13.16]	[-3.74]				
Outgroup x Trump voter	−0.60***	−0.08					−0.63***	−0.20					−0.63***	−0.15					−0.58***	−0.07				
	(0.16)	(0.13)					(0.15)	(0.13)					(0.15)	(0.13)					(0.16)	(0.13)				
	[-13.78]	[-1.92]					[-15.15]	[-4.87]					[-15.13]	[-3.69]					[-13.13]	[-1.76]				
SpiteQ	0.46***	0.11***	0.48***	0.10*	0.44***	0.12**													0.54***	0.13***	0.55***	0.10*	0.53***	0.16***
	(0.04)	(0.03)	(0.06)	(0.04)	(0.05)	(0.04)													(0.04)	(0.03)	(0.06)	(0.04)	(0.06)	(0.05)
Social Distance	−0.07·	−0.07*	−0.10	0.0003	−0.05	−0.11*													−0.09*	−0.07*	−0.13*	−0.003	−0.06	−0.12**
	(0.04)	(0.03)	(0.07)	(0.05)	(0.05)	(0.04)													(0.04)	(0.03)	(0.07)	(0.05)	(0.05)	(0.04)
Feeling Thermometer	0.02	0.02	−0.002	0.02	0.04	0.01													0.02	0.01	0.0002	0.02	0.03	0.02
	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)													(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)
Age							0.004	0.002	0.0001	0.01	0.01·	−0.0002							0.01***	0.004	0.01·	0.01*	0.01***	0.001
							(0.003)	(0.003)	(0.005)	(0.004)	(0.004)	(0.003)							(0.003)	(0.003)	(0.01)	(0.004)	(0.004)	(0.004)
EduLow							−0.12	−0.21**	−0.11	−0.27**	−0.11	−0.17·							−0.06	−0.23**	−0.04	−0.28**	−0.05	−0.18·
							(0.08)	(0.07)	(0.13)	(0.10)	(0.11)	(0.10)							(0.09)	(0.08)	(0.14)	(0.11)	(0.12)	(0.10)
IncomeLow							0.17*	0.16*	0.25*	0.13	0.11	0.18*							0.11	0.14*	0.18	0.07	0.08	0.18*
							(0.07)	(0.06)	(0.12)	(0.09)	(0.09)	(0.09)							(0.08)	(0.07)	(0.13)	(0.10)	(0.10)	(0.09)
Male							0.33***	0.14*	0.19·	0.02	0.42***	0.21*							0.51***	0.21**	0.39**	0.08	0.60***	0.29**
							(0.07)	(0.06)	(0.12)	(0.09)	(0.09)	(0.09)							(0.08)	(0.07)	(0.13)	(0.10)	(0.10)	(0.09)
PovertyCounty													0.02*	−0.01·	0.04**	0.004	0.01	−0.03**	0.02*	−0.01·	0.05**	0.01	0.01	−0.03**
													(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)
PovertyState													0.02	−0.03	−0.03	−0.02	0.06*	−0.03	0.02	−0.03	−0.04	−0.02	0.06*	−0.02
													(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.04)	(0.03)	(0.03)	(0.03)
Viol.Crimes													−44.20	−23.73	−141.97	62.74	8.73	−60.54	−47.50	−26.31	−181.90·	50.95	30.50	−61.05
													(54.34)	(47.74)	(89.69)	(73.16)	(69.30)	(62.76)	(56.66)	(47.48)	(94.36)	(73.46)	(72.42)	(62.33)
Crimes													−16.02	8.57	−4.82	9.19	−20.69	4.65	−18.48	6.47	−0.77	8.89	−28.66	4.20
													(16.46)	(14.20)	(26.10)	(21.12)	(21.38)	(19.02)	(17.10)	(14.15)	(27.20)	(21.18)	(22.36)	(18.94)
Religion													−1.00	1.55·	−2.45	2.07	0.25	0.95	−1.47	1.17	−2.99	1.90	−0.21	0.29
													(1.13)	(0.97)	(1.81)	(1.42)	(1.46)	(1.31)	(1.17)	(0.96)	(1.90)	(1.41)	(1.51)	(1.31)
BeliefInGod													1.84	−0.87	4.69	−3.48	−0.48	0.98	2.58	−0.02	5.91	−3.28	0.18	2.45
													(2.31)	(1.95)	(3.82)	(2.97)	(2.93)	(2.58)	(2.39)	(1.95)	(3.99)	(2.97)	(3.03)	(2.58)
Model	Logistic	Beta	Logistic	Beta	Logistic	Beta	Logistic	Beta	Logistic	Beta	Logistic	Beta	Logistic	Beta	Logistic	Beta	Logistic	Beta	Logistic	Beta	Logistic	Beta	Logistic	Beta
LogLik	-2000.07	87.54	-775.18	10.14	-1223.42	89.51	-2171.99	91.85	-850.38	10.71	-1319.31	95.02	-2065.04	85.83	-805.03	9.85	-1253.45	93.03	-1961.04	105.94	-758.04	18.88	-1192.38	109.4
Observations	3,054	1,342	1,174	564	1,880	778	3,221	1,415	1,234	596	1,987	819	3,054	1,342	1,174	564	1,880	778	3,054	1,342	1,174	564	1,880	778

Notes: ·p<0.10;*p<0.05;**p<0.01;***p<0.001;

Table 7: Zero-inflated beta regression of spite using socio-demographics and further controls.

Outgroup denotes a dummy with value one if the opponent is an outgroup-member (i.e., opposing voter). Trump voter denotes a dummy with value one if the deciding participant is a Trump voter and zero if the deciding participant is a Clinton voter. SpiteQ denotes the responses in the spite questionnaire. Social Distance denotes how close participants indicate to feel towards their opponent. Feeling Thermometer denotes the feeling of warmth participants indicate to feel towards their opponent. EduLow denotes participants who have at most a high school degree. IncomeLow denotes participants who have an income below 70k a year. PovertyCounty denotes the poverty level reported in (United States Department of Agriculture, 2017) on the county level. PovertyState denotes the poverty level reported in (Glassman and United States Census Bureau, 2016) on the state level. Viol.Crimes(County)/Crimes(County) denotes the percentage of (violent) crimes relative to the population of a county reported in (Federal Bureau of Investigation, United States Department of Justice, 2016). Religion denotes the percentage of people in a county indicating that religion is very important in their lives. BeliefInGod indicates the percentage of people in a county stating that they believe in God. Models denoted by "Spite?" estimate the decision to behave spitefully or not with a logistic regression. Models denoted by "Score" estimate the decision on how spitefully to behave conditionally on behaving spitefully using a beta regression. Standard errors are in parenthesis. Marginal effects (in %) are shown in brackets.

Comparison	DF	χ^2_{diff}	P-value
Beta (LogLik=83.898) vs. Exp. (LogLik=-363.643)	1.00	895.08	≤0.001***
Beta (LogLik=83.898) vs. Log-Normal (LogLik=-43.358)	0.00	254.51	≤0.001***
Beta (LogLik=83.898) vs. Gaus (LogLik=-18.506)	0.00	204.81	≤0.001***
Beta (LogLik=83.898) vs. Exp.Gaus (LogLik=20.563)	-1.00	126.67	≤0.001***
Beta (LogLik=83.898) vs. Gamma (LogLik=44.869)	0.00	78.06	≤0.001***
Beta (LogLik=83.898) vs. Weibul (LogLik=61.694)	0.00	44.41	≤0.001***

Table 8: Model fit of alternative models.
This table compares the Log-likelihoods of different models estimating the conditional spite score. Beta denotes a beta regression. Exp. denotes an exponential regression. Log-Normal denotes a log-normal regression. Gaus denotes a standard Gaussian regression. Exp.Gaus denotes an exponential gaussian regression. Gamma denotes a gamma regression and Weibul denotes a weibul regression. χ^2_{diff} denotes the Chi-squared test statistic of the differences.

C.6. Demand effects

As mentioned earlier, we made the manipulation of the opponent very salient. One possible concern the reader might have is a demand effect. To estimate the bounds of a possible demand effect in our setting, we applied the method suggested by de Quidt et al. (2018). More specifically, we conducted an experiment exactly as described in the section 2 with the exception that we did not provide participants with any information concerning their partner but rather told them "Note, that we expect that participants who have been shown our instructions (compared to a group with different instructions) will behave more/less spitefully in the subsequent tasks." Thus, we conducted one treatment to obtain the bound on the positive demand effect ("...more spitefully ...") and one treatment to obtain the bound on the negative demand effect ("...less spitefully ..."). We can compare these two treatments to the neutral baseline treatment, where there also was no information provided upon the partner. We conducted this experiment as part of the fifth wave in early January 2021. 171 participants finished the "negative demand treatment," and 165 participants finished the "positive demand treatment".

Figure 5 depicts the average spite behavior in the two demand treatments as well as the baseline treatment. The average spite score in the "negative demand treatment" is $M = 0.17$ ($SD = 0.29$), and it is $M = 0.17$ ($SD = 0.27$) in the "positive demand treatment" – both are statistically indistinguishable ($t(334)$= -0.0, $p\geq0.05$). The spite score in the baseline treatment is $M = 0.19$ ($SD = 0.28$), which also is indistinguishable from the negative demand treatment ($t(483)$= -0.6, $p\geq0.05$), the positive demand treatment ($t(477)$= -0.7, $p\geq0.05$), and both demand treatments pooled ($t(648)$= -0.8, $p\geq0.05$).[53] Thus, we see that even actively inducing

[53]Using more sophisticated estimations leads to the same results. Specifically, we estimate

a demand does not change the behavior.

Another concern a reader might have is that if there is a demand effect, this might affect Clinton and Trump voters differently. Specifically, it might be that Clinton voters react stronger to demand or social desirability, which might explain our heterogeneous result mentioned in section 3. However, we can see from Figure 5 that Clinton and Trump voters react to the same extent to the induced demand. More specifically, we find that the induced demand does not affect the behavior of either Clinton or Trump voters, nor do we find any interaction effect using linear regressions, Tobit regressions as well as zero-inflated beta-regressions. This additional insight thus shows that demand effects do not threaten our results.

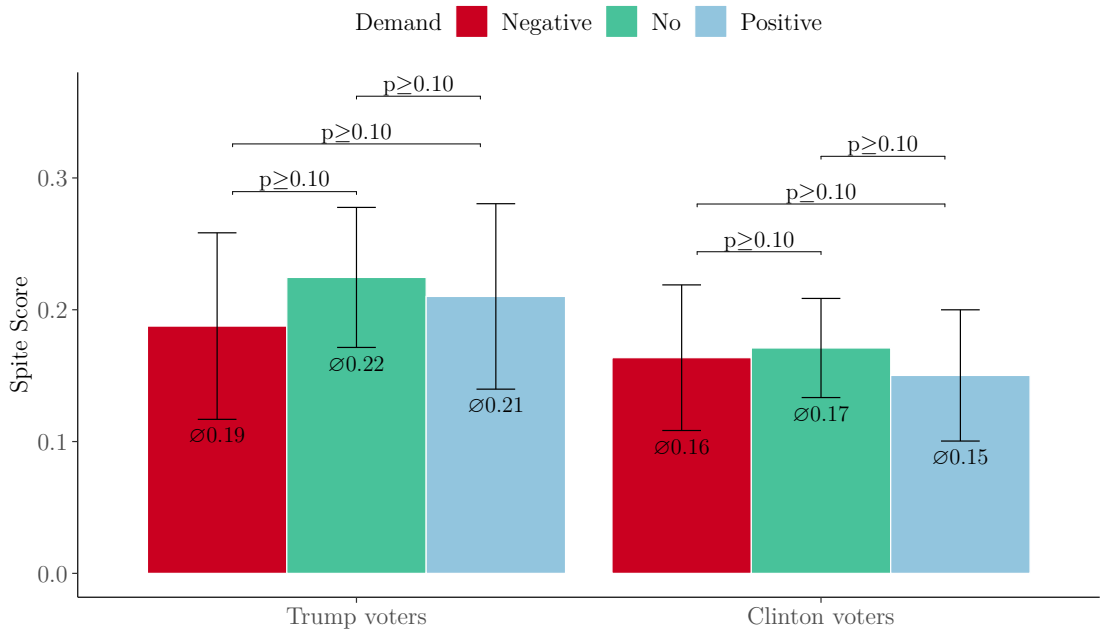


Figure 5: Spite by vote and demand

The figure depicts how spiteful participants behave towards their partners. The left three bars show the spite behavior of Trump voters while the right three bars show the spite behavior of Clinton voters. Red bars denote the spite behavior in the negative demand treatment, while blue bars denote the spite behavior in the positive demand treatment. Green bars denote the spite behavior in the baseline treatment without an induced demand effect. Tie fighters denote 95% confidence intervals. P-values are calculated using t-tests.

no significant difference between the two demand treatments among themselves as well as compared to the baseline using linear regressions, Tobit regressions as well as zero-inflated beta-regressions.

D. Alternative estimations

In this section, we will estimate the same models as in section 3.2. However, we will use rather ordinary tools to estimate the spite score. Hence, in this section, we will not use the zero-inflated beta regression, and therefore, we will ignore that the spite score is between 0 and 1.

If we do not account neither for the bounded outcome of the spite score nor for the zero inflation we could simply use a linear regression. The outcome is reported in Table 9 in Models (1)-(4). It shows the same estimation as Table 2 of section 3.2.

However, as it seems very reasonable to account for the extensive amount of zeros in the spite score, we use also a standard Tobit regression. Table 9 shows in models (5)-(8) the same estimation as Table 2 of section 3.2.

It can be seen from the results, that all main results still prevail using the ordinary estimation approach. In particular, we see, on average, more spiteful behavior towards outgroup-members than towards ingroup-members. Also, as before, we find that most of the spite behavior is driven by Clinton voters while Trump voters do not behave significantly more spitefully towards outgroup-members.

	Linear regressions				Tobit regressions			
	Full sample		Trump voters	Clinton voters	Full sample		Trump voters	Clinton voters
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Constant	0.19***	0.17***	0.24***	0.17***	−0.09***	−0.17***	0.04·	−0.19***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.03)	(0.03)
Outgroup	0.03**	0.06***	−0.02	0.06***	0.08***	0.16***	−0.05	0.16***
	(0.01)	(0.01)	(0.02)	(0.01)	(0.02)	(0.03)	(0.03)	(0.03)
Trump voter		0.07***				0.19***		
		(0.02)				(0.03)		
Trump voter x Outgroup		−0.09***				−0.20***		
		(0.02)				(0.05)		
Model	OLS	OLS	OLS	OLS	Tobit	Tobit	Tobit	Tobit
Observations	3221	3221	1234	1987	3221	3221	1234	1987
LogLik	-542.98	-531.11	-191.59	-339.34	-2361.61	-2346.06	-896.67	-1444.93
Notes:	·p<0.10;*p<0.05;**p<0.01;***p<0.001;							

Table 9: Estimating spite with ordinary methods.

Outgroup denotes a dummy with value one if the opponent is an outgroup-member (i.e. opposing voter). Trump voter denotes a dummy with value one if the deciding participant is a Trump voter and zero if the deciding participant is a Clinton voter. Models (1), (2), and (5),(6) report on the full sample while Models (3),(7), and (4),(8) report on Trump and Clinton voters, respectively. Models (1)-(4) show the estimations obtained using a linear regression. Estimations in Models (5)-(8) are performed using a Tobit regression. Standard errors are in parenthesis.

D.1. Propensity score matching

Even though participants were assigned an opponent randomly, participants did select whether to be a Trump or a Clinton voter. Throughout most of the paper, we only compare the behavior towards an opponent within a group (either Trump or Clinton voters), but in Table 2, we also compare whether Clinton and Trump voters differ in their spite behavior.

While we cannot simply randomly assign participants to be a Trump or a Clinton voter, we can match participants on *observabale* characteristics to control for self-selection. Specifically, we use nearest neighbor propensity score matching to estimate Tables 2, and 9. Moreover, we reestimate Tables 15, and 14. More specifically, we match participants on the following demographic characteristics: Age, gender, education, ethnicity, and income. Table 10 shows the estimation using propensity score matching. It is evident from Table 10 that all results are

qualitatively not influenced by participants self-selecting into Clinton and Trump voters.

	Spite								Thermo		Distance	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Constant	−0.10***	−0.19***	0.19***	0.15***	−0.42***	−0.69***	0.20***	0.20**	6.91***	6.93***	5.06***	5.09***
	(0.02)	(0.02)	(0.01)	(0.01)	(0.05)	(0.07)	(0.05)	(0.07)	(0.04)	(0.05)	(0.02)	(0.03)
Outgroup	0.09***	0.18***	0.03***	0.07***	0.30***	0.57***	−0.06	0.01	−5.59***	−6.03***	−2.24***	−2.59***
	(0.02)	(0.03)	(0.01)	(0.01)	(0.07)	(0.09)	(0.06)	(0.08)	(0.05)	(0.06)	(0.03)	(0.03)
					[7.31]	[13.88]	[-1.50]	[0.19]				
Trump voter		0.21***		0.09***		0.69***		0.01		−0.10		−0.08
		(0.03)		(0.02)		(0.11)		(0.10)		(0.07)		(0.04)
						[16.71]		[0.20]				
Outgroup x Trump voter		−0.22***		−0.09***		−0.70***		−0.18		1.15***		0.91***
		(0.04)		(0.02)		(0.15)		(0.13)		(0.10)		(0.05)
						[-16.86]		[-4.37]				
Model	Tobit	Tobit	OLS	OLS	Logistic	Logistic	Beta	Beta	Mixed effects	Mixed effects	Mixed effects	Mixed effects
LogLik	-2329.03	-2308.65	-1003.31	-986.93	-2364.3	-2343.33	35.88	37.97	-14785.35	-14686.93	-11003.16	-10796.22
AIC	4664.06	4627.3	2012.62	1983.86	4732.6	4694.66	-65.75	-65.93	29578.7	29385.86	22014.32	21604.44
Observations			3,221	3,221	3,221	3,221	1,415	1,415	6,442	6,442	6,442	6,442

Notes: *p<0.05;**p<0.01;***p<0.001;

Table 10: Reestimating results using propensity score matching.
Outgroup denotes a dummy with value one if the opponent is an outgroup-member (i.e., opponent voted not the same candidate on election day) and zero otherwise. Trump voter denotes a dummy with value one if the deciding participant is a Trump voter and zero if the deciding participant is a Clinton voter. The first two columns reestimate the spite behavior reported in Table 9 using a Tobit regression. Models (3) and (4) reestimate the spite behavior reported in Table 9 using a linear regression. Models (5) and (6) reestimate the spite behavior reported in Table 2 using a logistic regression on the decision to be spiteful or not. Models (7) and (8) reestimate the spite behavior reported in Table 2 using a beta regression on the conditional spite behavior. Models (9), (10) and (11), (12) reestimate the feeling of warmth reported in Table 15 and the social distance reported in Table 14 using mixed-effects models. Standard errors are in parenthesis. Marginal effects (in %) are shown in brackets were appropriate. All reestimations are performed using new weights obtained from nearest neighbor propensity score matching on the following demographics: Age, gender, education, ethnicity, and income.

D.2. Weighted sample

As pointed out in section B, our sample is not representative. Even though the assignment of the opponent is random, and all results can be interpreted causally, the result might not be representative of the US population. Thus, we reweight our sample to make the results more representative. Specifically, we change the weights of our sample by using the R `anesrake` method, which performs raking to obtain the weights such that our demographic variables coincide with the target demographic of the US population. The target demographic variables are: age, gender, ethnicity, and educational achievement.[54]

Using the obtained weights we estimate spite behavior reported in Tables 2, and

[54] Age and gender demographics can be found here: https://www.census.gov/prod/cen2010/briefs/c2010br-03.pdf. Information on educational achievement can be found here: https://www.census.gov/data/tables/2018/demo/education-attainment/cps-detailed-tables.html. The distribution of ethnicity can be found here: https://www.census.gov/quickfacts/fact/table/US/PST045218.

9. Moreover, we reestimate the attitudes reported in Tables 15, and 14. Table 11 shows the estimation. It is evident from Table 11 that all results are qualitatively not influenced by our sample not being representative.

	Spite		Spite		Spite		Spite		Thermo		Distance	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Constant	−0.11***	−0.21***	0.17***	0.14***	−0.37***	−0.69***	0.08	−0.06	6.83***	6.65***	4.96***	4.88***
	(0.02)	(0.02)	(0.01)	(0.01)	(0.05)	(0.06)	(0.04)	(0.06)	(0.04)	(0.05)	(0.02)	(0.03)
Outgroup	0.07**	0.14***	0.04***	0.07***	0.13	0.33***	0.09	0.33***	−5.32***	−5.36***	−2.01***	−2.17***
	(0.02)	(0.03)	(0.01)	(0.01)	(0.07)	(0.09)	(0.06)	(0.08)	(0.04)	(0.05)	(0.02)	(0.03)
					[3.12]	[7.87]	[2.15]	[8.14]				
Trump voter		0.33***		0.14***		1.16***		0.34***		0.46***		0.13**
		(0.03)		(0.01)		(0.11)		(0.09)		(0.08)		(0.05)
						[27.56]		[8.53]				
Outgroup x Trump voter		−0.26***		−0.12***		−0.81***		−0.61***		0.14		0.53***
		(0.05)		(0.02)		(0.16)		(0.13)		(0.09)		(0.05)
						[-19.19]		[-15.21]				
Model	Tobit	Tobit	OLS	OLS	Logistic	Logistic	Beta	Beta	Mixed effects	Mixed effects	Mixed effects	Mixed effects
LogLik	-2287.38	-2230.85	-2157.85	-2111.6	-2006.48	-1943.4	39.95	51.52	-16504.2	-16478.51	-12497.14	-12390.54
AIC	4580.75	4471.71	4321.69	4233.2	4016.96	3894.79	-73.89	-93.03	33016.4	32969.01	25002.28	24793.09
Observations			3,221	3,221	3,221	3,221	1,415	1,415	6,442	6,442	6,442	6,442

Notes: *p<0.05;**p<0.01;***p<0.001;

Table 11: Reestimating results using population weights.
Outgroup denotes a dummy with value one if the opponent is an outgroup-member (i.e., opponent voted not the same candidate on election day) and zero otherwise. Trump voter denotes a dummy with value one if the deciding participant is a Trump voter and zero if the deciding participant is a Clinton voter. The first two columns reestimate the spite behavior reported in Table 9 using a Tobit regression. Models (3) and (4) reestimate the spite behavior reported in Table 9 using a linear regression. Models (5) and (6) reestimate the spite behavior reported in Table 2 using a logistic regression on the decision to be spiteful or not. Models (7) and (8) reestimate the spite behavior reported in Table 2 using a beta regression on the conditional spite behavior. Models (9), (10) and (11), (12) reestimate the feeling of warmth reported in Table 15 and the social distance reported in Table 14 using mixed-effects models. Standard errors are in parenthesis. Marginal effects (in %) are shown in brackets were appropriate. All reestimations are performed using weights to make our sample more representative. The weights are obtained by using the R `anesrake` method, which performs raking to obtain the weights such that our demographic variables coincide with the target demographic of the US population (i.e., age, gender, ethnicity, and educational achievement).

E. Wave effects

As mentioned in the design section, we conducted the experiment in five waves: in late November 2016 (before the 58th US presidential election), late January 2017 (after the inauguration), late October 2018 (before the midterms), early November 2018 (after the midterms), and early January 2021 (after the 59th US presidential election). Even though neither spite nor attitudes change systematically over time (as can be seen in Figure 6), we account for the timing effects in the following subsections were we focus on each measure separately.

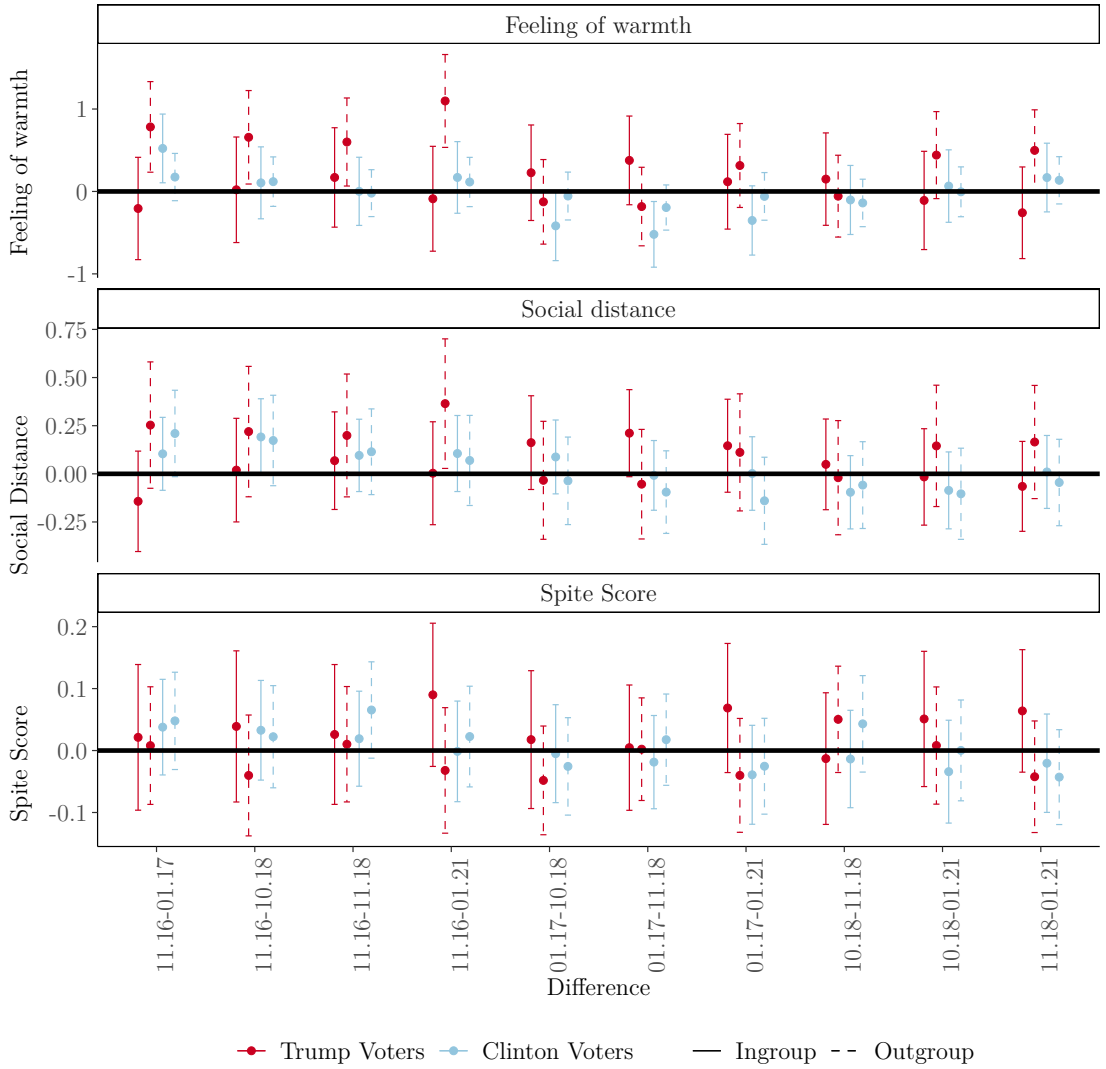


Figure 6: Differences in attitudes and spite over time

The figure depicts the pairwise difference in spite behavior and attitudes between each wave. The panel on the bottom represents the spite behavior. The panel in the middle represents the reported social distance. The panel on top depicts the reported feeling of warmth. Blue lines denote the behavior and attitudes of Clinton voters, while red lines represent the behavior and attitudes of Trump voters. Solid lines depict the behavior and attitudes towards ingroup-members (i.e., coinciding voter), while dashed lines depict the behavior and attitudes towards outgroup-members (i.e., opposing voter). The differences between waves are: from after the inauguration in late January 2017 to before the 58th US presidential election in late November 2016 (11.16-01.17); from before the midterms in late October 2018 to before the 58th US presidential election in late November 2016 (11.16-10.18); from after the midterms in early November 2018 to before the 58th US presidential election in late November 2016 (11.16-11.18); and from after 59th US presidential election in early January 2021 to before the 58th US presidential election in late November 2016 (11.16-11.18); All the other comparisons are denoted accordingly. Tie fighters depict 95% confidence intervals.

E.1. Time and Spite

In this subsection, we discuss how our results change if we take the timing into account and if we focus on the results for each wave individually.

Table 12 reports the estimation of a zero-inflated beta regression while controlling for the wave specific effects. In addition, Table 13 reports the estimation of a linear model and a Tobit model of spite while controlling for the wave specific effects. We can see that all results prevail if accounting for wave specific effects.

	Spite?	Spite Score	Spite?	Spite Score	Spite?	Spite Score	Spite?	Spite Score
	Full sample		Full sample		Trump voters		Clinton voters	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Constant	-0.37^{***}	0.34^{***}	-0.61^{***}	0.40^{***}	0.02	0.24^{*}	-0.61^{***}	0.41^{***}
	(0.09)	(0.08)	(0.10)	(0.09)	(0.14)	(0.11)	(0.11)	(0.10)
Outgroup	0.25^{***}	-0.04	0.50^{***}	0.02	-0.12	$-0.17^{\cdot}$	0.50^{***}	0.02
	(0.07)	(0.06)	(0.09)	(0.08)	(0.11)	(0.09)	(0.09)	(0.09)
	[6.20]	[-1.02]	[12.04]	[0.48]	[-3.03]	[-4.17]	[11.81]	[0.42]
Trump voter			0.62^{***}	-0.12				
			(0.11)	(0.10)				
			[15.08]	[-2.83]				
Outgroup x Trump voter			-0.62^{***}	-0.18				
			(0.15)	(0.13)				
			[-15.10]	[-4.53]				
Model	Logistic	Beta	Logistic	Beta	Logistic	Beta	Logistic	Beta
Wave fixed effects	✓	✓	✓	✓	✓	✓	✓	✓
Observations	3,221	1,415	3,221	1,415	1,234	596	1,987	819
Log Likelihood	−2,198.46	78.84	−2,182.42	85.93	−853.62	8.49	−1,327.41	91.42

Notes: $^{\cdot}$p<0.10;*p<0.05;**p<0.01;***p<0.001;

Table 12: Zero-inflated beta regression model of spite behavior.

Outgroup denotes a dummy with value one if the opponent is an outgroup-member (i.e. opposing voter). After election denotes a dummy with value one if the decision was made after the inauguration of Donald Trump and zero if the decision was made before the election of Donald Trump. Models (1), (3), and (5) estimate the decision to behave spitefully or not with a logistic regression. Models (2), (4), and (6) estimate the decision on how spitefully to behave conditionally on behaving spitefully using a beta regression. All estimations account for wave specific effects. The omitted category is after the inauguration. Standard errors are in parenthesis. Marginal effects (in %) are shown in brackets.

	Linear regressions				Tobit regressions			
	Full sample		Trump voters	Clinton voters	Full sample		Trump voters	Clinton voters
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Constant	0.20***	0.18***	0.24***	0.18***	−0.08**	−0.15***	0.05	−0.16***
	(0.01)	(0.01)	(0.02)	(0.02)	(0.03)	(0.03)	(0.04)	(0.04)
Outgroup	0.03**	0.06***	−0.02	0.06***	0.07***	0.16***	−0.04	0.16***
	(0.01)	(0.01)	(0.02)	(0.01)	(0.02)	(0.03)	(0.03)	(0.03)
Trump voter		0.07***				0.18***		
		(0.02)				(0.03)		
Trump voter x Outgroup		−0.09***				−0.20***		
		(0.02)				(0.05)		
Model	OLS	OLS	OLS	OLS	Tobit	Tobit	Tobit	Tobit
Wave fixed effects	✓	✓	✓	✓	✓	✓	✓	✓
Observations	3221	3221	1234	1987	3221	3221	1234	1987
LogLik	-539.51	-528.05	-190.53	-335.68	-2357.8	-2342.78	-895.93	-1440.84
Notes:	·p<0.10;*p<0.05;**p<0.01;***p<0.001;							

Table 13: Ordinary regressions of the spite score

Outgroup denotes a dummy with value one if the opponent is an outgroup-member (i.e. opposing voter). Trump voter denotes a dummy with value one if the deciding participant is a Trump voter and zero if the deciding participant is a Clinton voter. Models (1), (2), and (5),(6) report on the full sample while Models (3),(7), and (4),(8) report on Trump and Clinton voters, respectively. Models (1)-(4) show the estimations obtained using a linear regression. Estimations in Models (5)-(8) are performed using a Tobit regression. All estimations account for wave specific effects. The omitted category is after the inauguration. Standard errors are in parenthesis.

Moreover, Figure 7 depicts the spite behavior within each wave. It is evident that the patterns reported in section 3.2 can also be seen within each wave. Trump voters are seemingly indifferent between Clinton and Trump voters (with a small tendency towards more spite behavior towards Clinton voters, with the exception of the third and fifth wave); Clinton voters are less spiteful towards ingroup-members than are Trump voters and we see that Clinton voters significantly differentiate between ingroup-members and outgroup-members.

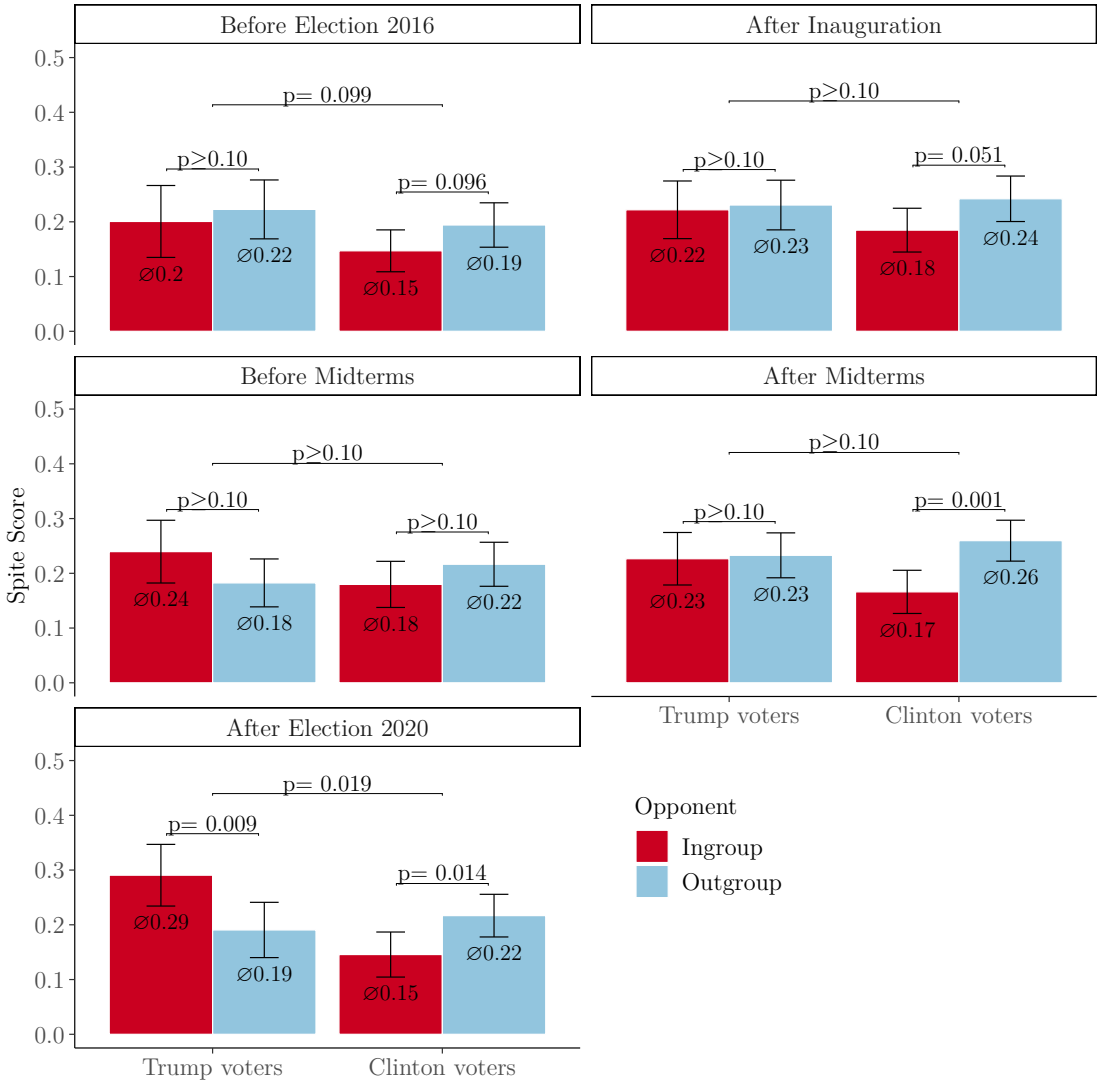


Figure 7: Spite towards opponents

The figure depicts how spiteful participants behave towards their ingroup-members (i.e., coinciding voter) and outgroup-members (i.e., opposing voter). The left two bars show the spite behavior of Trump voters while the right two bars show the spite behavior of Clinton voters. Red bars denote the spite behavior towards ingroup-members (i.e., coinciding voters) while blue bars denote the spite behavior towards outgroup-members (i.e., opposing voters). Tie fighters depict 95% confidence intervals. P-values are calculated using t-tests. The upper left panel shows attitudes before the 58th US presidential election. The upper right panel shows attitudes after the inauguration. The middle left and right panel depict the attitudes before and after the midterms, respectively. The bottom left panel shows attitudes after the 59th US presidential election.

E.2. Time and Attitdues

In this subsection, we focus on the attitudes participants reported to have in each of the five waves.

E.2.1. Time and feeling of warmth

Concerning the feeling of warmth, Table 14 reports the attitudes indicated by participants accounting for wave specific effects. We can see that accounting for wave specific effects does not change the results substantially and that all main results prevail. In addition, Figure 8 depicts the feeling of warmth in each wave. We can also see that, as reported before, typically Clinton and Trump voters do not differ in their attitudes towards ingroup-members; that Clinton and Trump voters have substantially and significantly less positive attitudes towards their opposing voters and that Clinton voters are also substantially and significantly more negative towards their outgroup-members than Trump voters are. We can see that the attitudes do vary slightly between waves, which, however, does not change the overall pattern nor any results.

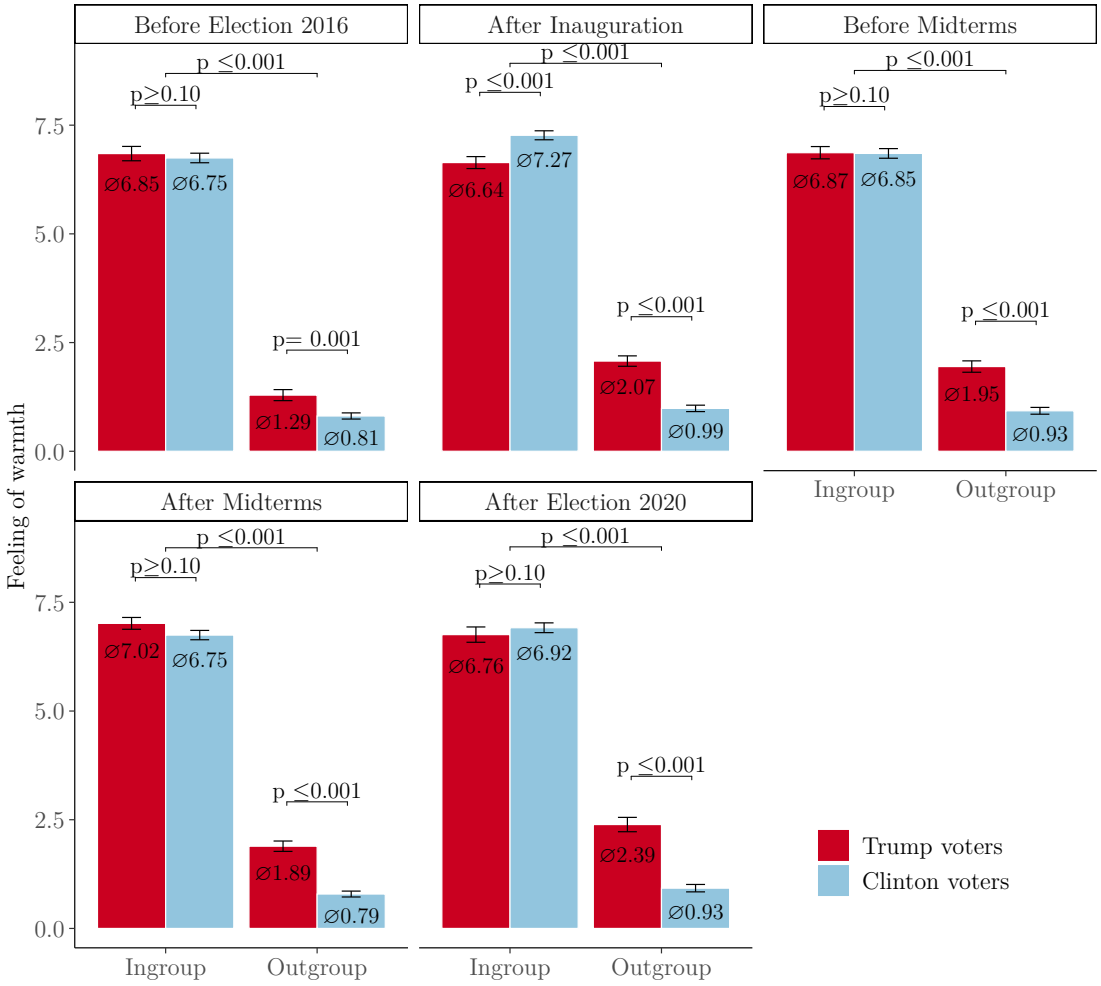


Figure 8: Feeling of warmth and opponent.

The figure depicts how warm participants report to feel towards their ingroup-members (i.e., coinciding voter) and outgroup-members (i.e., opposing voter). The left two columns show how warm participants felt towards their ingroup-members (i.e., coinciding voters) while the two right columns indicate how warm participants felt towards their outgroup-members (i.e., opposing voters). Red bars denote the attitudes of Trump voters, while blue bars denote the attitudes of Clinton voters. Tie fighters denote 95% confidence intervals. P-values are calculated using t-tests. The upper left panel shows attitudes before the 58th US presidential election. The upper right panel shows attitudes after the inauguration. The middle left and right panel depict the attitudes before and after the midterms, respectively. The bottom left panel shows attitudes after the 59th US presidential election.

	Difference in feelings towards ingroup- and outgroup-members		Feeling of warmth	
	(1)	(2)	(3)	(4)
Constant	6.07*** (0.06)	6.06*** (0.12)	0.89*** (0.05)	1.02*** (0.07)
Outgroup			6.02*** (0.06)	6.02*** (0.06)
Trump voter	−1.05*** (0.10)	−1.05*** (0.10)	1.06*** (0.07)	1.05*** (0.07)
Trump voter x Outgroup			−1.14*** (0.10)	−1.14*** (0.10)
Model	OLS	OLS	Mixed Effects	Mixed Effects
Wave fixed effects	×	✓	×	✓
Observations	3,221	3,221	6,442	6,442
Notes:		·p<0.10;*p<0.05;**p<0.01;***p<0.001;		

Table 14: Regression of feeling of warmth.
Outgroup denotes a dummy with value zero if the opponent is a ingroup-member and one if the opponent is an outgroup-member. Trump voter denotes a dummy with value one if the deciding participant is a Trump voter and zero if the deciding participant is a Clinton voter. Models (1) and (2) regress (using OLS) on the difference in the feeling of warmth towards ingroup-members and outgroup-members. Models (3) and (4) regresses simply on the feeling of warmth towards the opponent using a mixed-effects model. Models (2) and (4) account for wave specific effects. The omitted category is after the inauguration. Standard errors are in parenthesis.

E.2.2. Time and Social distance

Concerning the social distance Table, 15 reports the attitudes participants indicated to have accounting for wave specific effects. We can see that accounting for wave specific effects does not change the results substantially and that all main results prevail. In addition, Figure 9 depicts the reported social distance in each wave. We can also see that, as reported before, typically Clinton and Trump voters do not differ in their attitudes towards ingroup-members; that Clinton and Trump voters have substantially and significantly less positive attitudes towards their opposing voters and that Clinton voters are also substantially and significantly more negative towards their outgroup-members than Trump voters are. We can see that the attitudes do vary slightly between waves, which, however, does not change the overall pattern nor any results.

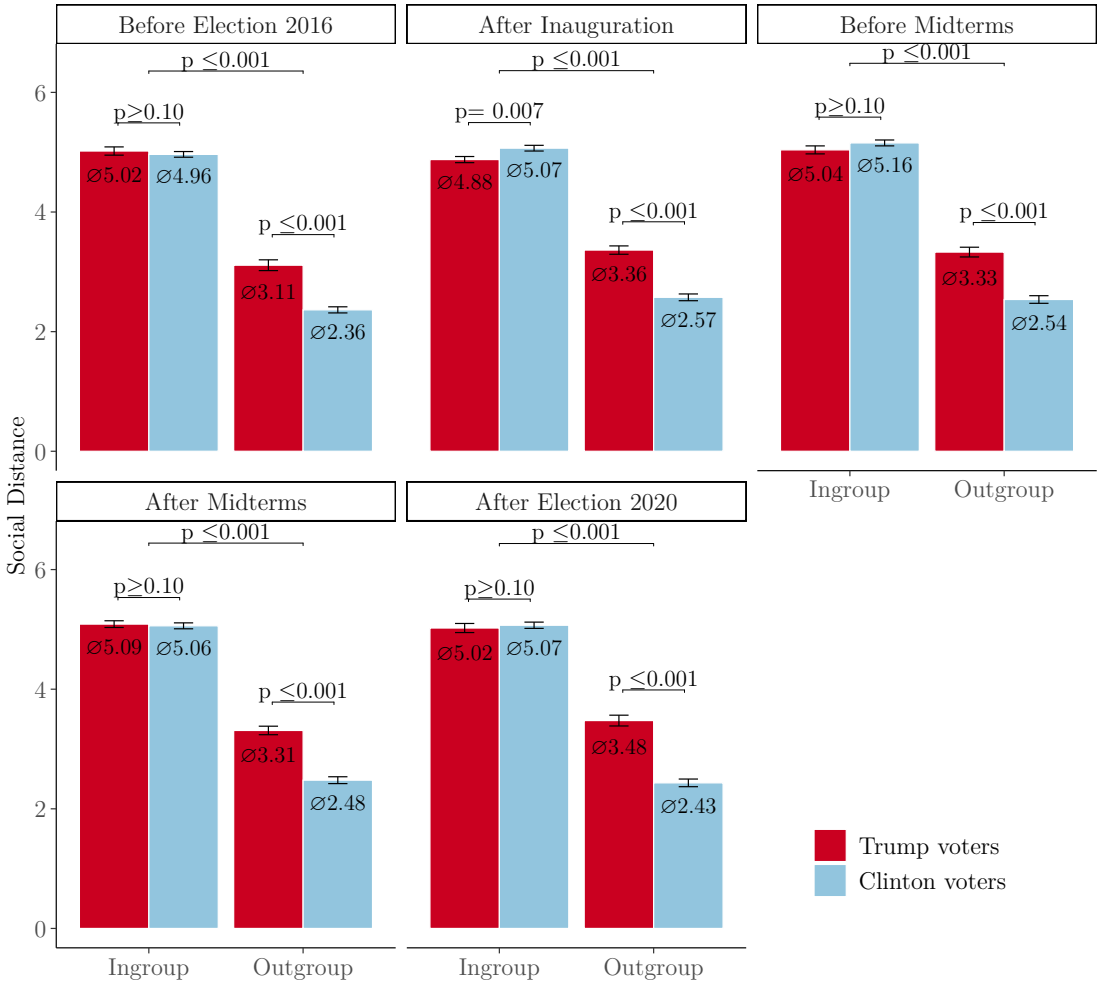


Figure 9: Social distance and opponent.

The figure depicts how close participants report to feel towards their ingroup-members (i.e., coinciding voter) and outgroup-members (i.e., opposing voter). The left two columns show how close participants felt towards their ingroup-members (i.e., coinciding voters) while the two right columns indicate how close participants felt towards their outgroup-members (i.e., opposing voters). Red bars denote the attitudes of Trump voters, while blue bars denote the attitudes of Clinton voters. Tie fighters denote 95% confidence intervals. P-values are calculated using t-tests. The upper left panel shows attitudes before the 58th US presidential election. The upper right panel shows attitudes after the inauguration. The middle left and right panel depict the attitudes before and after the midterms, respectively. The bottom left panel shows attitudes after the 59th US presidential election.

	Difference in social distance between ingroup- and outgroup-members		Social distance	
	(1)	(2)	(3)	(4)
Constant	2.61*** (0.04)	2.48*** (0.07)	2.48*** (0.02)	2.49*** (0.04)
Outgroup			2.58*** (0.03)	2.58*** (0.03)
Trump voter	−0.80*** (0.06)	−0.80*** (0.06)	0.85*** (0.04)	0.84*** (0.04)
Trump voter x Outgroup			−0.90*** (0.06)	−0.90*** (0.06)
Model	OLS	OLS	Mixed Effects	Mixed Effects
Wave fixed effects	×	✓	×	✓
Observations	3,221	3,221	6,442	6,442
Notes:		·p<0.10;*p<0.05;**p<0.01;***p<0.001;		

Table 15: Regression of the social distance.
Outgroup denotes a dummy with value zero if the opponent is a ingroup-member and one if the opponent is an outgroup-member. Trump voter denotes a dummy with value one if the deciding participant is a Trump voter and zero if the deciding participant is a Clinton voter. Models (1) and (2) regress (using OLS) on the difference in the social distance towards ingroup-members and outgroup-members. Models (3) and (4) regresses simply on the social distance towards the opponent using a mixed-effects model. Models (2) and (4) account for wave specific effects. The omitted category is after the inauguration. Standard errors are in parenthesis.

F. Path analysis

We have seen that participants' attitudes towards their opponent depend on whether the opponent is an ingroup-member or an outgroup-member (result A1). We also have seen that Clinton voters, exhibit stronger emotions towards their outgroup-members (result A2). Additionally, Clinton voters show more spite behavior towards outgroup-members than do Trump voters (result B1). Hence, it seems plausible that attitudes might mediate the decision to be spiteful.

To estimate the indirect relationship of the opponent for Trump voters and Clinton voters (see Figure 11 for the path-diagram) on the spite behavior via the attitude towards the opponent, we use a structural equation model (SEM), a common approach, especially in psychology. However, this approach is not very common in political science and economics, and hence, we discuss the results and the method only in the appendix.

First, we will compare several structural equation models (SEMs) and use the best to derive insight about the indirect effect of partisanship on behavior.

F.1. Model comparisons SEM

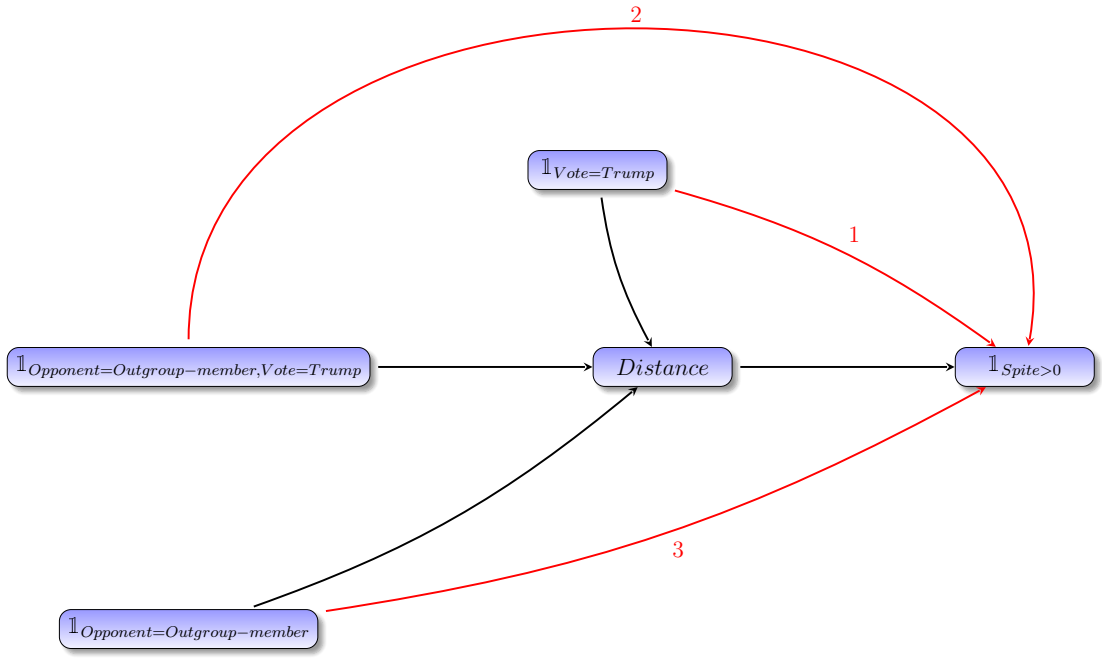


Figure 10: Possible structural equation models on the effect of the binary decision to behave spitefully or not, via social distance.

In this section, we investigate which structural model performs best in describing the effect on spite via social distance and also via feeling of warmth. As we were not able to find a canned version in R of a SEM with a zero-inflated beta regression, we used only the binary decision to be spiteful. A SEM with a standard OLS of the spite score leads to substantially identical results.

For that purpose, we compare the basic structural equation shown in Figure 10 as the model with black edges with extended models, including further direct effects. Red edges are extensions of the basic model, including, for example, the edge 1 denotes the direct effect of vote on spite. Table 16 shows the relevant model comparisons for the SEM via social distance and Table 17 shows the relevant model comparisons for the SEM via feeling of warmth. It can be seen that the best model is the full model (i.e., the basic model plus all the direct effect on spite) for a structural equation model via social distance as well as feeling of warmth.

Hence, the structural equation model we use has a direct and indirect effect of vote on the decision to behave spitefully at all and in addition, it has an indirect and direct effect of the opponent and a direct and indirect effect of the interaction of

opponent and vote on the decision to behave spitefully. The resulting model can be seen in Figure 11 for the social distance and Figure 12 for the feeling of warmth model.

Comparison	DF	Chisq diff	P-value
Model:Basic + 1 vs. Model:Basic	1	32.39	≤0.001***
Model:Basic + 2 vs. Model:Basic	1	13.82	≤0.001***
Model:Basic + 3 vs. Model:Basic	1	6.28	= 0.012*
Model:Basic + 1 + 2 vs. Model:Basic	2	38.50	≤0.001***
Model:Basic + 1 + 3 vs. Model:Basic	2	39.15	≤0.001***
Model:Basic + 2 + 3 vs. Model:Basic	2	19.53	≤0.001***
Model:Basic + 1 + 2 + 3 vs. Model:Basic	3	44.35	≤0.001***
Model:Basic + 1 + 3 vs. Model:Basic + 1	1	6.74	= 0.009**
Model:Basic + 1 + 2 vs. Model:Basic + 1	1	13.88	≤0.001***
Model:Basic + 1 + 2 +3 vs. Model:Basic + 1 + 2	1	8.45	= 0.004**

Table 16: Comparing the possible structural equation models on the effect of spite via social distance.

The basic model is a model with direct effects of vote, opponent, and the interaction of opponent and vote on the social distance plus a direct effect of social distance on spite. Model "Basic + 1" is the basic model plus a direct effect of vote on spite etc.

Comparison	DF	Chisq diff	P-value
Model:Basic + 1 vs. Model:Basic	1	32.91	≤0.001***
Model:Basic + 2 vs. Model:Basic	1	16.29	≤0.001***
Model:Basic + 3 vs. Model:Basic	1	9.91	= 0.002**
Model:Basic + 1 + 2 vs. Model:Basic	2	40.43	≤0.001***
Model:Basic + 1 + 3 vs. Model:Basic	2	42.55	≤0.001***
Model:Basic + 2 + 3 vs. Model:Basic	2	25.66	≤0.001***
Model:Basic + 1 + 2 + 3 vs. Model:Basic	3	49.90	≤0.001***
Model:Basic + 1 + 3 vs. Model:Basic + 1	1	9.95	= 0.002**
Model:Basic + 1 + 2 vs. Model:Basic + 1	1	16.30	≤0.001***
Model:Basic + 1 + 2 +3 vs. Model:Basic + 1 + 2	1	11.58	= 0.001***

Table 17: Comparing the possible structural equation models on the effect of spite via feeling of warmth.

The basic model is a model with direct effects of vote, opponent, and the interaction of opponent and vote on the feeling of warmth plus a direct effect of the feeling of warmth on the decision to behave spitefully. Model "Basic + 1" is the basic model plus a direct effect of vote on spite etc.

F.2. Results of the SEM

F.2.1. SEM and social distance

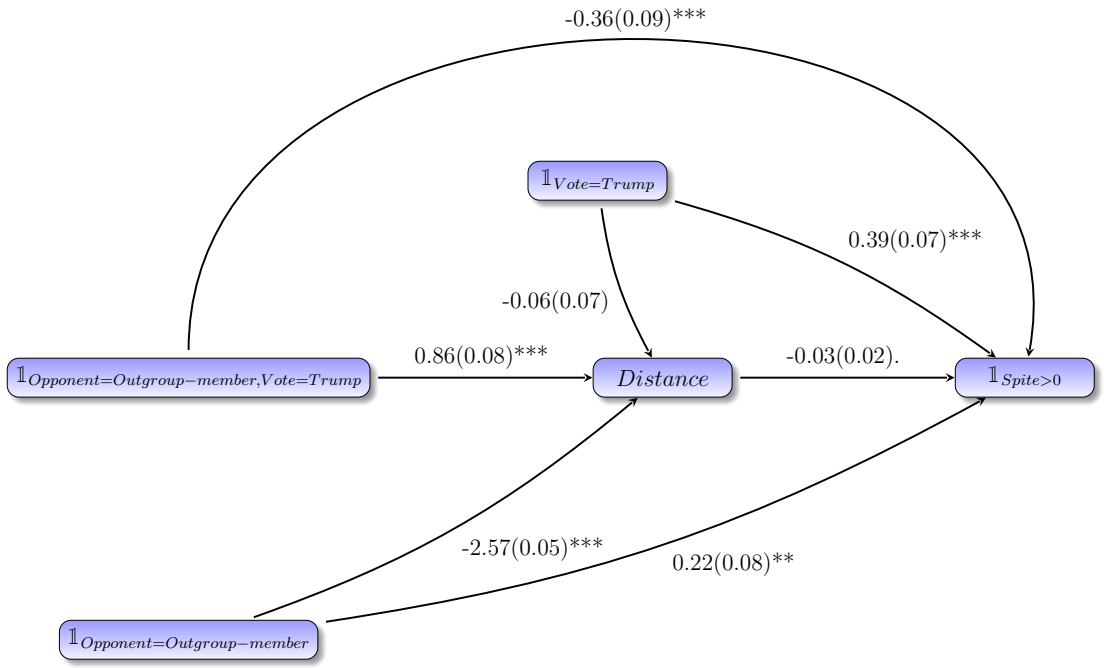


Figure 11: SEM of spite via social distance.
Structural Equation Model on the effect on the binary decision to behave spitefully or not, via social distance. Numbers in brackets denote standard errors.

As Figure 11 shows and as seen before, the opponent and the interaction of Trump voters and opponent had a statistically significant influence on the social distance ($p_{\text{Opponent}} \leq 0.001$, $p_{\text{Interaction}} \leq 0.001$) while Trump voters did not show significantly different attitudes in the social distance towards ingroup-members compared to Clinton voters ($p_{\text{Vote}} \geq 0.10$).

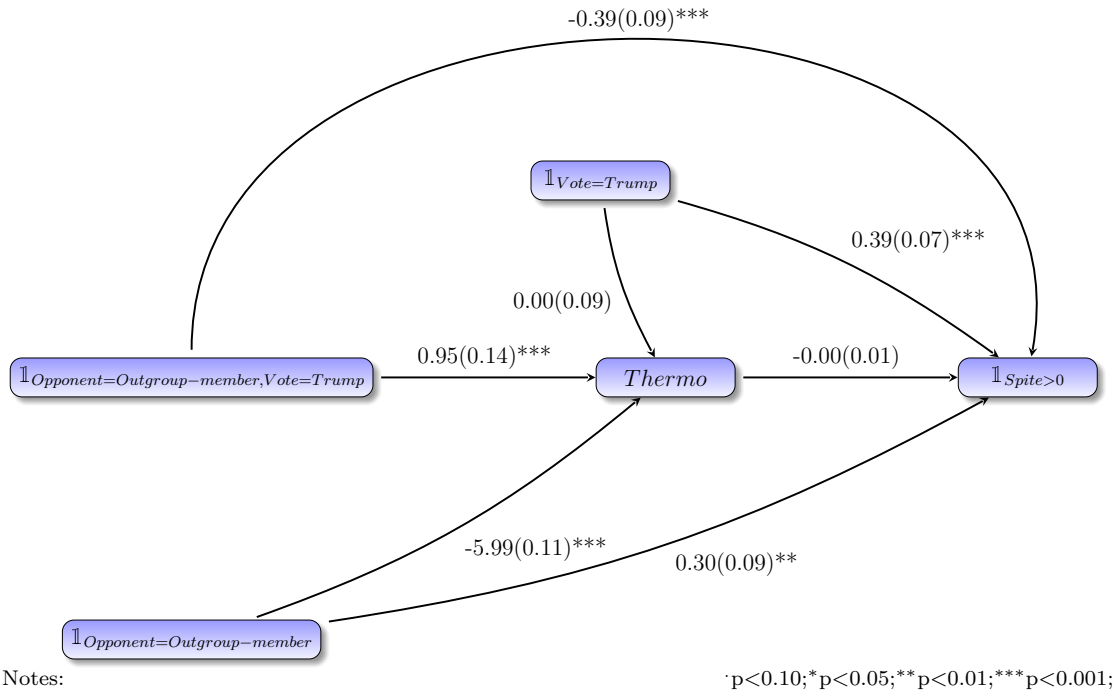
Additionally, all direct effects statistically significantly influence the decision to behave spitefully ($p_{\text{Distance}} = 0.097$, $p_{\text{Vote.Direct}} \leq 0.001$, $p_{\text{Interaction.Direct}} \leq 0.001$, $p_{\text{Opponent.Direct}} = 0.004$).

Using 10,000 bootstrapped samples we obtain the 95% confidence intervals for the indirect effects of the opponent ([-0.016,0.186]), the vote ([-0.003,0.007]) and the interaction of opponent and vote ([-0.063,0.006]). As the confidence intervals for the indirect effects all include the zero, we cannot reject the null hypothesis of no indirect effect. However, the total effects (combining both the indirect and direct effects) are statistically significant for the opponent ([0.198,0.422]), the vote

([0.259,0.527]), and the interaction of opponent and vote ([-0.568,-0.207]).

Thus, we have seen that Trump voters have a higher odd of behaving spitefully, that Clinton voters have a higher odd of behaving spitefully towards their outgroup-members, that Trump voters have a lower odd of behaving spitefully towards their outgroup-members than Clinton voters, and that an increased feeling of closeness leads to a decrease in the odds of behaving spitefully. We also have seen that the opponent strongly influences the feeling of closeness and that Clinton and Trump voters differ in their attitudes towards their opponents. Overall, we can see that Trump voters have higher odd of behaving spitefully, which is not mediated through the social distance, and that Clinton voters have a higher odd of behaving spitefully towards their outgroup-members which is not mediated through social distance and that Trump voters have a lower odd of behaving spitefully towards their outgroup-members than Clinton voters which also is not mediated through social distance.

F.2.2. SEM and feeling of warmth



Notes: $^{\cdot}$p<0.10;*p<0.05;**p<0.01;***p<0.001;

Figure 12: SEM of spite via feeling of warmth.
Structural Equation Model on the effect on the binary decision to behave spitefully or not, via the feeling of warmth. Numbers in brackets denote standard errors.

Using the same model for the indirect effect of feeling of warmth leads to similar results. The opponent and the interaction of Trump voters and opponent had a statistically significant influence on the social distance ($p_{\text{Opponent}} \leq 0.001$, $p_{\text{Interaction}} \leq 0.001$) while Trump voters did not show significantly different attitudes in the social distance towards ingroup-members compared to Clinton voters ($p_{\text{Vote}} \geq 0.10$).

Additionally, all direct effects (with the exception of the direct effect of feeling of warmth) statistically significantly influence the decision to behave spitefully ($p_{Distance} \geq 0.10$, $p_{\text{Vote.Direct}} \leq 0.001$, $p_{\text{Interaction.Direct}} \leq 0.001$).

Using again 10,000 bootstrapped samples we obtain the 95% confidence intervals for the indirect effects of the opponent ([-0.123,0.141]), the vote ([-0.000,0.000]) and the interaction of opponent and vote ([-0.022,0.020]). As with the social distance model, we cannot reject the null hypothesis of no indirect effects. However, the total effects (combining both the indirect and direct effects) are statistically significant for the opponent ([0.198,0.422]), the vote ([0.259,0.527]), and the interaction of opponent and vote ([-0.568,-0.207]).

Thus, we have seen that Trump voters have a higher odd of behaving spitefully, and that Trump voters have a lower odd of behaving spitefully towards their outgroup-members than Clinton voters. We also have seen that the opponent strongly influences the feeling of warmth and that Clinton and Trump voters differ in their attitudes towards their opponents. Overall, we can see that Trump voters have higher odd of behaving spitefully, which is not mediated through the social distance, and that Clinton voters have a higher odd of behaving spitefully towards their outgroup-members which is not mediated through social distance and that Trump voters have a lower odd of behaving spitefully towards their outgroup-members than Clinton voters which also is not mediated through social distance.

Thus, we obtain basically the same results if feeling of warmth is used as the measure of attitudes instead of the feeling of closeness.

G. Spite questionnaire

To have a better understanding of the general spite tendencies of participants, we elicited the spitefulness of participants by using the questionnaire by Marcus et al. (2014), which is based on 17 items. Examples of the 17 questions are the following:[55]

- If I am checking out at a store and I feel like the person in line behind me is rushing me, then I will sometimes slow down and take extra time to pay.

[55] All questions are shown in related materials.

- I would rather no one get extra credit in a class if it meant that others would receive more credit than me.

Participants were asked to indicate their agreement on a scale between 1 and 5. Higher scores on the scale indicate higher spitefulness.

Figure 13 shows the classification of how spiteful participants are over time for Clinton and Trump voters. We can see that Clinton and Trump voters do not differ substantially[56] and that there is no substantial change over time in attitudes. As the measure does not rate the spite *towards* others but rather the individual spite attitude there are no differences in spite attitudes between those participants who have been assigned an outgroup-member as the opponent and those who have been assigned an ingroup-member.[57] Given that there are no substantial differences between Clinton and Trump voters, the results are omitted to the appendix and are reported solely for reasons of transparency.

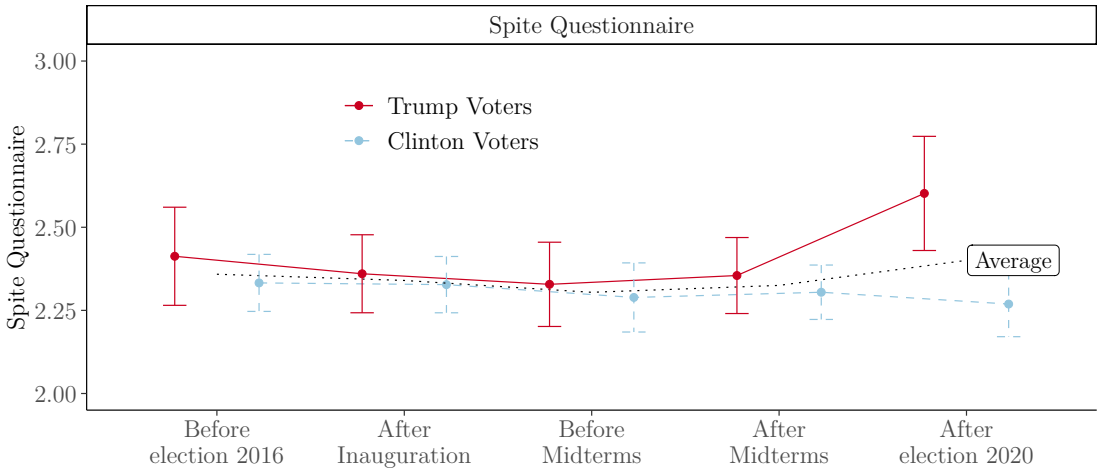


Figure 13: Spite questionnaire
The figure depicts how spiteful participants were classified following the spite questionnaire in each of the five waves. Red solid lines denote the general spite attitudes of Trump voters while blue dashed lines denote the general spite attitudes of Clinton voters. The black dotted line depicts the average spite attitude. Tie fighters denote 95% confidence intervals.

[56] Trump voters seem to be classified as slightly more spiteful (M = 2.41, SD = 1.08) than Clinton voters (M = 2.31, SD = 0.92), $t(3219)$= -2.9, p= 0.004.

[57] Ingroup-members: M = 2.36, SD = 1; Outgroup-members M = 2.33, SD = 0.96. The difference is not significant, $t(3219)$= 0.7, $p \geq 0.05$.

H. Moral attitudes

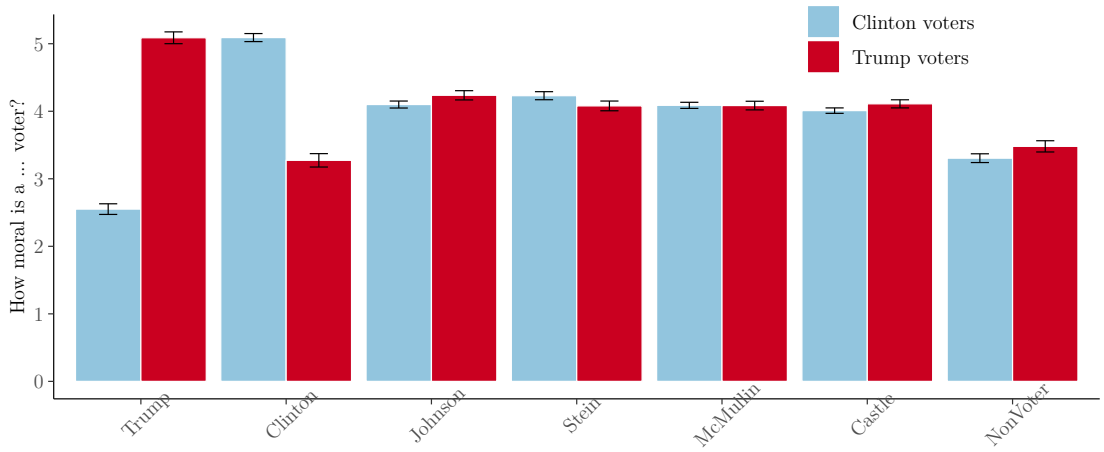


Figure 14: How moral participants report to consider different voters.
The figure depicts how moral participants consider Trump, Clinton, Johnson, Stein, McMullin, and Castle voters as well as people who did not vote during the presidential election of 2016. The scale goes from 1 (not moral at all) to 7 (very moral). Red bars denote the attitudes of Trump voters, while blue bars denote the attitudes of Clinton voters. Tie fighters denote 95% confidence intervals.

To further investigate whether the perception of morality might drive the difference between Clinton and Trump voters, we elicit their moral attitudes during the third, fourth and fifth wave of the experiment (late October 2018, early November 2018, early January 2021). In particular, we asked participants how moral they consider a Trump, Clinton, Johnson, Stein, McMullin, and Castle voter as well as people who did not vote during the presidential election 2016. The results are shown in Figure 14.

The moral attitudes of both Clinton and Trump voters are very similar and rather neutral for Johnson, Stein, McMullin, and Castle voters. Both Clinton and Trump voters considered non-voters significantly less moral than Johnson, Stein, McMullin, and Castle voters.[58] Interestingly, Clinton voters considered Clinton voters as moral as Trump voters considered Trump voters[59], which was significantly better than Johnson, Stein, McMullin, and Castle voters ($t(1958)$= -31.7, $p \leq 0.001$).

[58] In particular, non-voters were considered to be $M = 3.38$ points moral on a scale from one to seven while Johnson, Stein, McMullin, and Castle voters were considered on average to be $M = 4.11$ points moral on a scale from one to seven, a highly significant difference ($t(1958)$= 23.1, $p \leq 0.001$).

[59] In particular, Clinton voters considered fellow Clinton voters to be $M = 5.09$ points moral on a scale from one to seven while Trump voters considered fellow Trump voters to be $M = 5.09$ points moral on a scale from one to seven, $t(1957)$= 0.1, $p \geq 0.05$.

More importantly, we can see that Trump voters considered Clinton voters significantly less moral than fellow Trump voters[60], than Johnson, Stein, McMullin, Castle voters[61] and even less than non-voters[62].

The same pattern can be found for Clinton voters who considered Trump voters significantly less moral than fellow Clinton voters[63], than Johnson, Stein, McMullin, Castle voters[64] and even less than non-voters[65]. Even more interestingly, the difference in morality between ingroup-members and outgroup-members is much more pronounced for Clinton voters, who considered Trump voters significantly less moral than Trump voters considered Clinton voters.[66]

Table 18 also estimates how the attitudes towards the morality of the opponent affect spiteful behavior. We see that a more positive evaluation of the opponent's morality substantially reduces spiteful behavior. We find this effect for both Clinton and Trump voters. Notably, a positive evaluation of the opponent's morality (i.e. values above 5) can fully remove the additional spiteful behavior of Clinton voters towards Trump voters.

The results provide suggestive evidence that morality might, in fact, be driving the difference in attitudes of participants and, in turn, might influence spite behavior.

[60] In particular, Trump voters considered fellow Trump voters to be $M = 5.09$ points moral on a scale from one to seven while Clinton voters were considered to be $M = 3.27$ points moral on a scale from one to seven, a highly significant difference $t(782) = 24.9$, $p \leq 0.001$.

[61] In particular, Trump voters considered Johnson, Stein, McMullin, Castle on average to be $M = 4.13$ points moral on a scale from one to seven while Clinton voters were considered to be $M = 3.27$ points moral on a scale from one to seven, a highly significant difference $t(782) = 14.5$, $p \leq 0.001$.

[62] In particular, Trump voters considered non-voters on average to be $M = 3.48$ points moral on a scale from one to seven while Clinton voters were considered to be $M = 3.27$ points moral on a scale from one to seven, a highly significant difference $t(782) = -3.8$, $p \leq 0.001$.

[63] In particular, Clinton voters considered fellow Clinton voters to be $M = 5.09$ points moral on a scale from one to seven while Trump voters were considered to be $M = 2.55$ points moral on a scale from one to seven, a highly significant difference $t(1175) = -44.3$, $p \leq 0.001$.

[64] In particular, Clinton voters considered Johnson, Stein, McMullin, Castle on average to be $M = 4.11$ points moral on a scale from one to seven while Trump voters were considered to be $M = 2.55$ points moral on a scale from one to seven, a highly significant difference $t(1175) = 34.1$, $p \leq 0.001$.

[65] In particular, Clinton voters considered non-voters on average to be $M = 3.31$ points moral on a scale from one to seven while Clinton voters were considered to be $M = 2.55$ points moral on a scale from one to seven, a highly significant difference $t(1175) = -17.7$, $p \leq 0.001$.

[66] In particular, Clinton voters considered Trump voters to be $M = 2.55$ points moral on a scale from one to seven while Trump voters considered Clinton voters to be $M = 3.27$ points moral on a scale from one to seven, a highly significant difference $t(1957) = -11.3$, $p \leq 0.001$.

	Spite?	Spite Score	Spite?	Spite Score	Spite?	Spite Score	Spite?	Spite Score
	Full sample		Full sample		Trump voters		Clinton voters	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Constant	−0.61*	0.05	−1.26**	0.57	0.14	−0.29	−1.26**	0.53
	(0.31)	(0.25)	(0.45)	(0.42)	(0.43)	(0.30)	(0.45)	(0.43)
Outgroup	1.10**	0.50·	1.70***	0.16	0.44	0.47	1.70***	0.15
	(0.34)	(0.28)	(0.48)	(0.44)	(0.50)	(0.35)	(0.48)	(0.45)
	[26.75]	[12.17]	[40.78]	[3.98]	[10.80]	[11.74]	[40.25]	[3.69]
Moral	0.06	0.05	0.13	−0.03	−0.01	0.10·	0.13	−0.03
	(0.06)	(0.05)	(0.09)	(0.08)	(0.08)	(0.06)	(0.09)	(0.08)
	[1.36]	[1.16]	[3.01]	[-0.73]	[-0.21]	[2.42]	[2.97]	[-0.67]
Outgroup x Moral	−0.26***	−0.20***	−0.31**	−0.18·	−0.23*	−0.15·	−0.31**	−0.17·
	(0.07)	(0.06)	(0.10)	(0.09)	(0.11)	(0.08)	(0.10)	(0.10)
	[-6.43]	[-4.88]	[-7.40]	[-4.35]	[-5.69]	[-3.79]	[-7.31]	[-4.01]
Trump voter			1.40*	−0.83				
			(0.62)	(0.53)				
			[33.61]	[-20.30]				
Outgroup x Trump voter			−1.26·	0.26				
			(0.69)	(0.58)				
			[-30.18]	[6.43]				
Trump voter x Moral			−0.13	0.12				
			(0.12)	(0.10)				
			[-3.21]	[2.88]				
Outgroup x Trump voter x Moral			0.08	0.04				
			(0.15)	(0.12)				
			[1.82]	[0.99]				
Model	Logistic	Beta	Logistic	Beta	Logistic	Beta	Logistic	Beta
Observations	1,959	882	1,959	882	783	379	1,176	503
Log Likelihood	−1,333.57	47.27	−1,319.35	52.64	−534.74	5.34	−784.61	56.85
Notes:			·p<0.10;*p<0.05;**p<0.01;***p<0.001;					

Table 18: Estimation of the spite behavior as a function of moral attitudes.
This table depicts a zero-inflated beta regression model of the spite behavior. Outgroup denotes a dummy with value one if the opponent is an outgroup-member (i.e., opposing voter) and zero otherwise. Trump voter denotes a dummy with value one if the deciding participant is a Trump voter and zero if the deciding participant is a Clinton voter. Moral denotes the moral attitudes (on a scale from 0 to 7) towards the opponent. Models (1), (3), (5), and (7) denoted by "Spite?" estimate the decision to behave spitefully or not with a logistic regression. Models (2), (4), (6), and (8) denoted by "Spite Score" estimate the decision on how spitefully to behave conditionally on behaving spitefully using a beta regression. Model (1)-(4) display the full sample while Models (5)-(6) and (7)-(8) display the spite behavior of Trump and Clinton voters, respectively. Standard errors are in parenthesis. Marginal effects (in %) are shown in brackets.

I. Altruism Score

As the focus of this paper is antisocial behavior, we mainly concentrate on our spite measure. However, we also collected data on prosocial behavior in all five

waves of our experiment. Specifically, participants had to make three distribution-decisions upon money in our spite measure (as described in section 2.3), and further, they needed to make six additional distribution-decisions upon money in the primary scale of the social value orientation (SVO) measure by Murphy et al. (2011). The primary scale of the SVO-task consists of six distribution-decisions among nine possible allocations. Table 19 shows these six decisions with all the possible allocations per decision. As we vary the identity of the receiver, we cannot rely on the original scoring by Murphy et al. (2011) which used the SVO-task as a measure of personality.[67] Instead, to obtain an altruism score, we use the procedure from the spite score. Specifically, the altruism score is the amount given to the opponent relative to the maximally possible amount. The amount given to the opponent can range between 0 and 290 points, and, therefore, the altruism score ranges between 0 and 1.

You receive	85	85	85	85	85	85	85	85	85
	○	○	○	○	○	○	○	○	○
Other receives	85	76	68	59	50	41	32	24	15
You receive	85	87	89	91	92	94	96	98	100
	○	○	○	○	○	○	○	○	○
Other receives	15	19	24	28	32	37	41	46	50
You receive	50	54	59	63	68	72	76	81	85
	○	○	○	○	○	○	○	○	○
Other receives	100	98	96	94	92	91	89	87	85
You receive	50	54	59	63	68	72	76	81	85
	○	○	○	○	○	○	○	○	○
Other receives	100	89	79	68	58	47	36	26	15
You receive	100	94	88	81	75	69	62	56	50
	○	○	○	○	○	○	○	○	○
Other receives	50	56	62	69	75	81	88	94	100
You receive	100	98	96	94	92	91	89	87	85
	○	○	○	○	○	○	○	○	○
Other receives	50	54	59	63	68	72	76	81	85

Table 19: Altruism measure.
The table depicts the nine allocation choices in each of the six decisions of participants in the primary scale of the social value orientation (SVO) measure by Murphy et al. (2011). For each choice, the upper row denotes the payoff in experimental currency units for the deciding participants, while the bottom rows each denote the payoff for the other player.

To alleviate possible concerns of a spite-demand effect (i.e., spiteful behavior being the only option to express the attitudes towards the opponent), we hid the spite-task within the SVO-task. Specifically, participants had to make all nine

[67] Nevertheless, using the original scoring by Murphy et al. (2011) would provide qualitatively the same results.

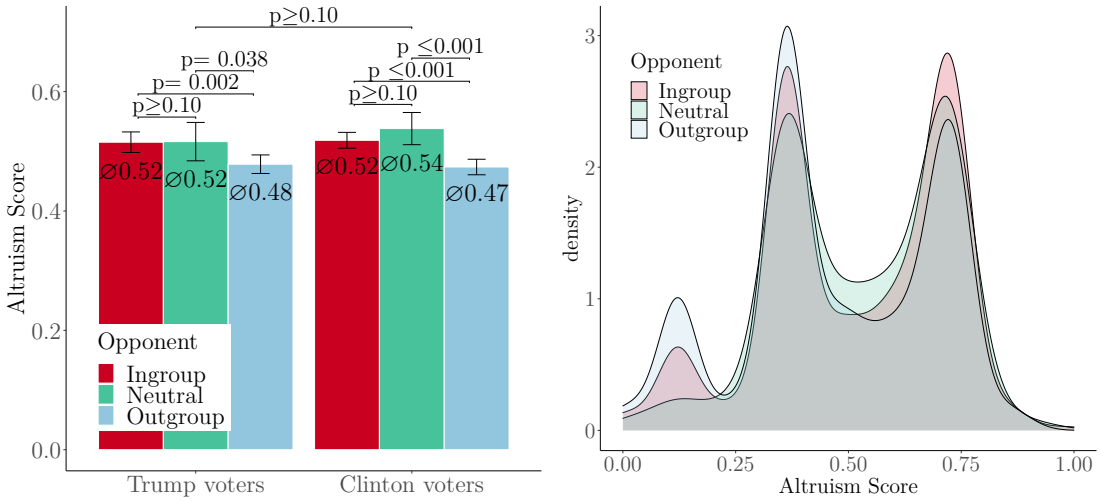
distribution-decisions (three from the spite-task and six from the SVO-task) on one screen, and all nine distribution-decisions have been ordered randomly. Thus, it is rather unlikely that participants chose to behave spitefully due to a lack of an alternative behavior as participants were presented with spiteful as well as altruistic options. Further, we find no evidence that the order of the presented options affects behavior. Specifically, we see that spite behavior is not affected by whether any spite-task-decision has been shown first or not. This indicates that the decisions most likely have been made conscious and not random.[68] Further evidence that the behavior has been conscious and not random is that we find clear treatment effects between people who have been assigned an ingroup-member as the opponent compared to people who have been assigned an outgroup-member. Figure 15b reports upon the distribution of the altruism score by opponent, and Figure 15a shows the average altruism score by opponent and vote.

It can be seen that, on average, ingroup-members were given more ($M = 0.52$; $SD = 0.21$) compared to outgroup-members ($M = 0.48$; $SD = 0.21$), $t(3154)$= 5.6, $p \leq 0.001$, which is in line with the findings typically reported in the literature on ingroup-outgroup behavior. However, we do not find ingroup-favoritism as ingroup-members were given roughly as much as neutral opponents ($t(491.1)$= -1.1, $p \geq 0.05$). We further do not see any difference between Clinton and Trump voters, neither in the behavior towards ingroup-members ($t(1159)$= 0.3, $p \geq 0.05$), outgroup-members ($t(1499.4)$= -0.5, $p \geq 0.05$), nor in their baseline behavior ($t(266.6)$= 1.0, $p \geq 0.05$).[69] This is very different to the results obtained from the spite behavior as well as the reports on attitudes. Thus, while Clinton and Trump voters omit help (i.e., how much is given towards the opponent) towards outgroup-members to the same extend, we find a substantial heterogeneous effect for spite behavior.

As a conclusion, we find conceptually similar patterns in behavior in altruistic and spiteful behavior. Interestingly, we see a heterogeneous effect for spite behavior between Clinton and Trump voters, which we do not see in altruistic behavior. This suggests that spite behavior is conceptually different from altruistic behavior.

[68] If randomness would be of relevance, we might expect that decisions made earlier are more accurate and that decisions made later are more random. Thus, we would expect a different behavior if one of the three spite measures were presented first. We find no difference between cases where spite was first ($M = 0.22$; $SD = 0.30$) and last ($M = 0.21$; $SD = 0.28$), $t(964.9)$= 1.5, $p \geq 0.05$. We also do not find any such order effects using more sophisticated models like linear regression, Tobit regressions as well as zero-inflated beta-regressions. Note further that we also do not find any interaction effect between the order and political affiliation. See section C.2.1 for a further discussion of noise being the driver of spiteful behavior.

[69] Using more sophisticated estimations leads to the same conclusion. Specifically, we estimate no significant interaction between Clinton and Trump voters using linear regressions, Tobit regressions as well as zero-inflated beta-regressions.



(a) Altruism towards opponents

(b) Distribution of the altruism score by opponent.

Figure 15: Results of the altruism score.

The figure on the left depicts how altruistically participants behave towards their ingroup-members (i.e., coinciding voter), outgroup-members (i.e., opposing voter) and neutral opponents. The left three bars show the altruism score of Trump voters while the right three bars show the altruism score of Clinton voters. Red bars denote the altruistic behavior towards outgroup-members, while blue bars denote the altruistic behavior towards ingroup-members averaged over all five waves. Green bars denote the altruistic behavior in the baseline treatment, which was collected only in the fifth wave, where no information upon the opponent was provided. Tie fighters depict 95% confidence intervals. P-values are calculated using t-tests.

The figure to the right depicts the distribution of the altruism score by opponent. The red distribution denotes the altruistic behavior toward ingroup-members, the blue distribution denotes the altruistic behavior towards outgroup-members and the green denotes the altruistic behavior in the baseline.

J. Instructions

In this section, we show the instructions for the own spite measure, the spite questionnaire, the feeling thermometer and the social distance questionnaire. Further, we show the manipulation and the attention check. The procedure of the experiment can be seen in Figure 16.

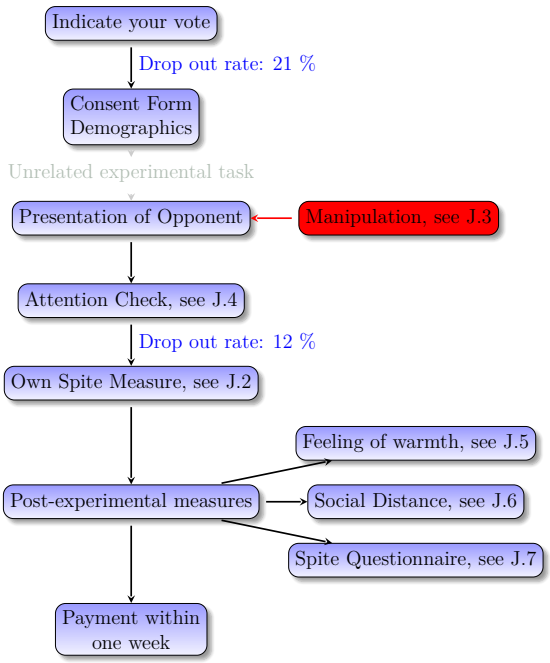


Figure 16: Procedure of the experiment.
The figure shows the procedure of the experiment.

J.1. Welcome

Welcome to this experiment in the economics of market decision making. If you follow these instructions carefully and make good decisions you will earn a considerable amount of money that will be paid to you within one week to your MTurk account. We ask that you pay close attention to the instructions.

Note that one of the main guidelines in the experimental economics is that we do NOT deceive participants (see for example https://en.wikipedia.org/wiki/Experimental_economics). Hence, all rules and restrictions will indeed be implemented in the way we describe them. We go to great lengths to ensure that assignments, randomization of variables and rules are implemented exactly in the

way they are presented here to you!

In this experiment, you will be assigned an opponent. Your payoff will depend on his/her decisions and his/her payoff may depend on your decisions. Typically every person is assigned, one opponent.

To comply with the non-deception-rules of economics we also need to inform you about a technical issue: It may happen that more than one person is assigned to another person. In such a (rather rare) case it will be randomly decided whose decision will be payoff relevant to this other person. Thus, your payoff will always depend on the decision of somebody else. Your decision will influence the payoff of your assigned partner in most cases. It, however, may happen that your decision does not impact the payoff of your partner as somebody else's choice has been determined payoff-relevant for your partner.

J.2. Own spite measure

In this task, you will be paired with another player, whom we will refer to as the opponent. All of your choices will be confidential. After you take your decisions this task will not be repeated and there is no further interaction with your opponent.

You will be making a series of decisions about allocating resources between you and your opponent. For each of the following questions, please indicate the distribution you prefer most by selecting the button below the payoff allocations. You can only make one selection for each question. Your decisions will yield money for both yourself and your opponent.

Each point shown is worth 0.2 cents (100 points = 20 cents).

In the example below, a person has chosen to distribute the payoff so that he/she receives 50 points (=10 cents), while his opponent receives 40 points (=8 cents).

There are no right or wrong answers, this is all about personal preferences. After you have made your decision, select the resulting distribution of money by clicking on the button below your choice. As you can see, your choices will influence both

the amount of money you receive as well as the amount of money your opponent receives.

At the end of the experiment, a computer program will randomly pick either you or your opponent as the payoff-relevant decision maker.

Only one of the following decisions will be payoff relevant. Which decision will be paid will be determined by a random process at the end of the experiment. Hence, you have to take all decisions seriously as any of those can be chosen by the random process with equal probability.

Please indicate your choice for each of the following distributions.
Note: These decisions are payoff relevant and will influence your payment!

[[Participants had to make choices as shown in Table 1]]

J.3. Assignment

[[Treatment 1:]]
Individuals who have indicated to vote for Donald Trump at the beginning of the experiment were assigned to a group called "red".
Individuals who have indicated to vote for Hillary Clinton at the beginning of the experiment were assigned to a group called "blue".

Your assigned opponent indicated to vote for Hillary Clinton. Hence, your opponent was assigned to be a member of the group "blue".

[[Positive demand treatment:]]
Note, that we expect that participants who have been shown our instructions (compared to a group with different instructions) will behave more spitefully in the subsequent tasks.

[[Treatment 2:]]
Individuals who have indicated to vote for Donald Trump at the beginning of the experiment were assigned to a group called "red".
Individuals who have indicated to vote for Hillary Clinton at the beginning of the experiment were assigned to a group called "blue".

Your assigned opponent indicated to vote for Donald Trump. Hence, your opponent was assigned to be a member of the group "red".

[[Negative demand treatment:]]
Note, that we expect that participants who have been shown our instructions (compared to a group with different instructions) will behave less spitefully in the subsequent tasks.

J.4. Attention Check

Which group does your opponent belong to? [[This was not shown in the baseline and demand treatments]]

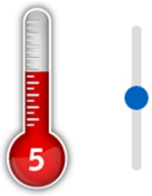
- (a) "Red"
- (b) "Blue"
- (c) "Green"

J.5. Feeling thermometer

This measure is called a "feeling thermometer" as it measures your feeling towards groups.

Here is how it works:
If you do not know too much about a group, or don't feel particularly warm or cold toward them, then you should place them in the middle, at the 5-degree mark.
If you have a warm feeling toward a group or feel favorably toward it, you would give it a score somewhere between 5 and 10 depending on how warm your feeling is toward the group.
On the other hand, if you don't feel very favorable toward some of these groups -if there are some you don't care for too much - then you would place them somewhere between 0 and 5.
Please indicate below your feeling towards the named group of people.

Where would you put the [[republicans]][[democrats]][[people who will vote for Hillary Clinton]][[people who will vote for Donald Trump]] on the thermometer?



J.6. Social distance questionnaire

Please rate the following statements. The person in question is a person who has indicated to vote for [[Hillary Clinton]][[Donald Trump]].

	Strongly agree	Agree	Somewhat agree	Neither agree nor disagree	Somewhat disagree	Disagree	Strongly disagree
This appears to be a likeable person.	○	○	○	○	○	○	○
I would like this person to be a close personal friend.	○	○	○	○	○	○	○
I would like this person to move into my neighborhood.	○	○	○	○	○	○	○
I would like this person to come and work at the same place I do.	○	○	○	○	○	○	○
This is a person who is similar to me.	○	○	○	○	○	○	○
I would like to have this person marry into my family.	○	○	○	○	○	○	○
This is the kind of person that I tend to avoid.	○	○	○	○	○	○	○

J.7. Spite questionnaire

The questions of the questionnaire according to Marcus et al. (2014) included the following questions:

- I would be willing to take a punch if it meant that someone I did not like would receive two punches.
- I would be willing to pay more for some goods and services if other people I did not like had to pay even more.
- If I was one of the last students in a classroom taking an exam and I noticed that the instructor looked impatient, I would be sure to take my time finishing the exam just to irritate him or her.
- If my neighbor complained about the appearance of my front yard, I would be tempted to make it look worse just to annoy him or her.
- It might be worth risking my reputation in order to spread gossip about someone I did not like.
- If I am going to my car in a crowded parking lot and it appears that another driver wants my parking space, then I will make sure to take my time pulling out of the parking space.
- I hope that elected officials are successful in their efforts to improve my community even if I opposed their election. (reverse scored)
- If my neighbor complained that I was playing my music too loud, then I might turn up the music even louder just to irritate him or her, even if meant I could get fined.
- I would be happy receiving extra credit in a class even if other students received more points than me. (reverse scored)
- Part of me enjoys seeing the people I do not like fail even if their failure hurts me in some way.
- If I am checking out at a store and I feel like the person in line behind me is rushing me, then I will sometimes slow down and take extra time to pay.
- It is sometimes worth a little suffering on my part to see others receive the punishment they deserve.
- I would take on extra work at my job if it meant that one of my co-

workers who I did not like would also have to do extra work.

- If I had the opportunity, then I would gladly pay a small sum of money to see a classmate who I do not like fail his or her final exam.
- There have been times when I was willing to suffer some small harm so that I could punish someone else who deserved it.
- I would rather no one get extra credit in a class if it meant that others would receive more credit than me.
- If I opposed the election of an official, then I would be glad to see him or her fail even if their failure hurt my community.