

A Partisan Treatment in a High Salience Election: Evidence from a Field Experiment in Germany *

MARCEL NEUNHOEFFER

How do partisan campaigns influence voting behavior in high salience elections? In 2016, I conducted a field experiment in the German state of Baden-Württemberg to evaluate the effectiveness of direct mail sent out by the Green party in a high salience election. I use a matched-pair randomized design to test the effects of a partisan valence direct mail treatment. Assessing the results with design-based methods of statistical inference, I find no effect of the letters on either turnout or vote choice. Nevertheless, this null result has important implications. First, political campaigns using direct mail as described in this study can save significant amounts of money. Second, this study shows that further experimental research within partisan campaigns is needed to better understand the effects of real campaign tools, especially in multi-party systems and high-salience elections.

How do partisan campaigns influence voting behavior in high salience elections? Despite the large number of get-out-the-vote field experiments (e.g., Gerber and Green 2000; Druckman et al. 2011; Nickerson 2008; Davenport 2010; Michelson 2003; Fieldhouse et al. 2014; Gerber, Green, and Larimer 2008; Panagopoulos 2010; Gerber and Green 2001; Dale and Strauss 2009; Gerber, Karlan, and Bergan 2009; Green and Gerber 2015), Michelson and Nickerson (2011) note that “[t]o date, only one high-profile campaign, that of Rick Perry in the 2006 Texas gubernatorial race, has agreed to participate

*I thank the Green Party of Germany (Bündnis 90/Die Grünen) to implement and fund the study in the state of Baden-Württemberg. I especially thank Sebastian Duwe, Robert Heinrich and Matthias Gauger for making this study possible. I received valuable comments from Florian Foos, Richard Traunmüller, Thomas Gschwend and Sebastian Reinkunz. Thank you very much. Furthermore, I thank the participants of the Causal Inference course at the CDSS in Mannheim for their helpful comments. To view supplementary material for this article, please visit xxx.

in a nonproprietary experimental study” (Michelson and Nickerson 2011, 235)¹. However, for the external validity of field experiments, especially when parties want to make inferences based on field experimental research, it is crucial to experiment within actual campaigns. “If well-funded and highly salient campaigns behave differently and/or voters respond differently to outreach from brand name organizations, then external validity is a real concern for much of the mobilization literature” (235).

Since 2011, one of the biggest partisan field experiments was conducted by Liegy, Muller, and Pons (2013) in France². They test the effect of partisan door-to-door canvassing in a nationwide experiment. In total, activists of the PS/Hollande campaign knocked on about 5 million doors, some of which were located in randomly assigned precincts. Pons (2016) estimates that the door-to-door campaign increased the vote share of Hollande and the candidates of PS by almost 2 percentage points.

Further partisan experiments were conducted in low saliency elections like municipal elections in the US (door-to-door and direct mail, Barton, Castillo, and Petrie 2014), a mayoral campaign in an Italian town (phone and direct mail, Kendall, Nannicini, and Trebbi 2015), a local and European Parliament election in the UK (leaflets and door-to-door, Foos and John 2016) and elections for a police and crime commissioner in the UK (phone, Foos and Rooij 2017).

With this paper I shed some light on the effects of partisan direct mail in a high salience election. The study is implemented in a hotly contested state election in Germany of national interest (elections in two other states were held at the same day). In the previous election average turnout was 66.3%. I had the unique opportunity to randomize the direct mail for the Green party’s campaign for the 2016 Baden-Württemberg state elections. In total, 82 small municipalities were selected to be part of the field experiment. Letters were sent out to about 23,000 households.

The contribution of this study is twofold. First, I contribute to the experimental literature by conducting a large scale field experiment within a partisan campaign, extending previous experimental research to a new context. To my knowledge, my study is the first assessing the effects of partisan direct mail in Germany. Furthermore, only

¹Results of the experimental evaluation of the Rick Perry campaign are reported in Gerber et al. (2011) and Shaw and Gimpel (2012).

²The results are also reported in Pons (2016).

few scientists had the opportunity to randomize actual campaigning efforts and just a handful of political scientists implemented field experimental research in Germany in general. Second, by experimenting within an actual campaign my contribution extends to campaign practitioners. I test an actual campaign method within an actual campaign.

I proceed as follows. First, I derive my theoretical expectations from a rational choice model of turnout and vote choice. Second, I describe my field experimental research design in detail. Third, I analyze the results of the field experiment. I conclude by discussing the implications of my null results for political science research as well as political election campaigns.

A SIMPLE MODEL OF TURNOUT AND VOTE CHOICE

Before turning to my experimental design, I need to state my expectations about how campaign communications – and my experimental treatment in particular – affect voting behavior. Let me start with a very simple theoretical intuition: Parties campaign to win votes. They can do so by either clarifying their ideological position or by making statements about the quality of their candidates³. Riker and Ordeshook (1968) formalize this intuition in their rational choice model of voting. Their model describes voting behavior as a two stage process. First, a voter decides whether to turnout or to abstain. Second, when going to the polls, a voter decides which option to vote for.

The simple intuition behind the first stage of the model is that voters vote when the benefit of voting exceeds the cost of voting. To formalize the model of Riker and Ordeshook (1968), they introduce four parameters: p , B_i , D_i and C_i . This can be expressed with Equation 1 for voter i :

$$pB_i + D_i > C_i, \tag{1}$$

where p is the probability that the vote will be decisive⁴ in bringing about the benefit B_i .

³The negative version of both could also be possible: Diffusing an opponent's position or make negative statements about the quality of opponent candidates. This is usually called negative campaigning. But in this paper I only consider positive campaigning.

⁴Estimates of p by Gelman, King, and Boscardin (1998) and Gelman, Silver, and Edlin (2012) for US presidential elections showed that p is about 1 in 10 million or 0.0000001 in states with a close election race and on average 1 in 60 million or 0.000000016 for voters

B_i , thus, is the benefit an individual voter receives when his or her preferred candidate prevails. D_i is the duty to vote a voter satisfies when she votes. Finally, C_i denotes the cost of voting⁵. As p is very small, especially for elections with millions of eligible voters, the decision to turnout is mainly driven by D_i and C_i . Although the original model was developed to describe two party or two candidate competition, equation 1 also applies to multi-party systems.

It is, however, more complicated to generalize stage two – the vote choice stage – of the model beyond the two party case. To still make use of the model, I assume that from a party's point of view, voters also face a binary choice problem: If a voter decides to turnout, she can either vote or not vote for the party under consideration. The vote choice is represented by B_i from Equation 1. Borrowing from Ashworth and Bueno de Mesquita (2009), I define B_i as:

$$B_i = \max_{j \in J} E(v_j - (x_i - x_j)^2). \quad (2)$$

In Equation 2, v_j denotes the valence of candidate or choice j . I define valence in the words of Stokes (1963), who calls “*valence-issues* those that merely involve the linking of the parties with some condition that is positively or negatively valued by the electorate” (373). x_i and x_j are the unidimensional ideal points⁶ of voter i and candidate or party j .

In the two party case, with $J = 2$ and $j = 1, 2$, a voter votes for party 1 if the following conditions are satisfied:

$$\begin{aligned} E(v_1 - (x_i - x_1)^2) &> E(v_2 - (x_i - x_2)^2) \\ \text{and} \\ pB_i + D_i &> C_i, \text{ with } B_i = E(v_1 - (x_i - x_1)^2). \end{aligned}$$

In the multiparty case, from party 1's point of view, $J = n$ and $j = 1, M$, with
in the US.

⁵Costs of voting usually include information costs and physical costs. Information costs could be, for example, the time spent on reading campaign material or looking for information when and where to vote. Physical costs of voting include getting to the polling place or waiting in line at the polling place.

⁶The most common ideological scale for those ideal points is a left-right scale.

$M = 2, \dots, n$ a voter votes for party 1, when

$$E(v_1 - (x_i - x_1)^2) > \max_{m \in M} E(v_m - (x_i - x_m)^2)$$

and

$$pB_i + D_i > C_i, \text{ with } B_i = E(v_1 - (x_i - x_1)^2).$$

Thus, the benefit of voting for party 1 has to be larger than the benefit of voting for the next best party. Partisan campaigns try to influence both stages. The effects of a campaign on stage two are rather intuitive. To win more votes, party 1 attempts to make itself more appealing to voters over the course of a campaign. They can either do so by trying to increase their valence, v_1 , or by positioning the party so that the ideological distance $(x_i - x_1)^2$ gets smaller for a larger number of voters.

The theoretical effect of party 1's campaign on the first stage in the model, turnout, however is more obscure. On the one hand, a partisan campaign usually contains some information about the date of an election or even the polling place; it can be assumed that C_i decreases. On the other hand, the direction of the effect of a partisan campaign on D_i is theoretically unclear (see Barton, Castillo, and Petrie 2014)⁷. The empirical evidence of partisan appeals on turnout suggests that at least on average there is no effect of partisan campaigns on turnout (see Pons 2016; Barton, Castillo, and Petrie 2014; Shaw and Gimpel 2012; Cardy 2005). Looking at the individual effects, Foos and Rooij (2017) and Foos and John (2016) find that partisan treatments increase turnout among voters who support the party that reached out to the voters, but not among supporters of other parties.

This leads me to the following expectations for a positive partisan valence treatment:

- I expect no effect of a positive partisan valence treatment on average turnout. If true, this would lend additional support to the theory (and previous empirical evidence) that partisan appeals – at least on average – do not affect citizens' duty to vote.

⁷For example, a partisan campaign treatment could increase the D_i of voters who would have voted for party 1 anyways, as they might feel a duty to help their party win. At the same time it could decrease the D_i of voters who would not have voted for party 1. The influence on D_i could also work the other way round. Or it could decrease D_i for all voters, or increase D_i for all voters.

- I further expect a positive effect of the partisan valence treatment on the party's vote share. If this is true, my experiment would provide empirical evidence that partisan valence appeals can influence the valence perception of voters, and subsequently their vote choice.

EXPERIMENTAL DESIGN

To test my expectations, I implement a field experiment in collaboration with the Green Party of Baden-Württemberg within their campaign for the state elections on March 13, 2016. For the experiment, I selected 82 small municipalities. 41 of those municipalities were randomly assigned to a partisan valence treatment by mail. In total, I sent out 22,662 letters to all households in the treatment group, and therefore all eligible voters in the treatment municipalities.

In this section, I first present the context of the campaign. Second, I describe my partisan valence treatment. Then, I briefly outline an ideal experimental design, before explaining crucial design adaptations I had to make due to several constraints. To close this section, I provide an overview of the election results in the state, as well as in my treatment and control municipalities.

Baden-Württemberg 2016 - A Green State and an Open Race

Baden-Württemberg is the third largest of the German federal states, both in area and population. In total about 10.8 million people live in Baden-Württemberg. For the elections on March 13, 2016 there were 7,685,778 citizens eligible to vote.

The German states have legislative power over education – in both schools and universities, police and some legislative power regarding infrastructure, among other policy domains. In 2016 the state had a budget of about 44 billion Euros. Furthermore, the federal state governments can also directly influence federal law-making through the second chamber, the Bundesrat.

The Green party went into the campaign with the incumbent minister-president Winfried Kretschmann, who led a green-red (Green Party and Social Democrats) coalition since the previous elections in 2011, as their lead candidate. Before the surprising win of the new coalition in 2011, the centre-right Christian Democratic Union (CDU) held the

minister-president office for 58 years and also was the strongest party in parliament ever since the foundation of the federal state in 1952. While they still remained the strongest party in parliament in 2011, the loss of the minister-president office in 2011 was a huge and unexpected defeat for the CDU. As expected, in 2016 Kretschmann's main challenger was a CDU candidate. The CDU chose Guido Wolf, former president of the state parliament and chairman of the CDU faction in the state parliament, as their lead candidate.

The state elections determine the composition of the state parliament. Subsequently, the minister-president has to be elected by a majority of the members of parliament. Members of parliament are elected for a term of five years. The term of the minister-president and the government usually corresponds to the term of the state parliament.

For state elections Baden-Württemberg is divided into 70 electoral districts. In contrast to federal elections in Germany, each voter has only one vote. On the ballot paper voters find the name of their district candidate and the candidate's party affiliation. The parliament has at least 120 members, 70 of whom are the respective winners of the 70 electoral districts. The total party seat share in parliament is determined according to the relative party results in the whole state with a seat distribution according to the Sainte-Laguë method. The polls leading up to the 2016 elections showed very high favorability ratings for Winfried Kretschmann. A representative survey leading up to the elections asked people who they preferred as their minister-president: more than 60% of the people wanted Winfried Kretschmann as the minister-president, while only 17% preferred Guido Wolf (see Forschungsgruppe Wahlen 2016). Nevertheless, in polls the Green Party was well behind the CDU at the start of the campaign, and the majority for the continuation of the green-red coalition was out of sight. However, the gap between the CDU and the Green Party was closing as election day was coming closer.

From a political point of view the race was an open one. Clearly, the main campaign goal of the Green Party was to keep Winfried Kretschmann in power as minister-president. The coalition partner, the Social Democrats, however, had bad poll results. That meant that the Greens had to try to become the strongest party in the state parliament. To that end, the whole campaign focused on Winfried Kretschmann and his qualities, trying to exploit his popularity across party boundaries. Similarly, the campaign of the CDU focused on their lead candidate Guido Wolf, despite his low favorability ratings.

According to the press agency *dpa* the Green party spent about 1,000,000 Euros on their campaign. The CDU spent 2,500,000 Euros and the SPD 2,200,000 (see Breining

2016). In general political campaigns in Germany are rather low-tech. The campaigns in Baden-Württemberg mostly spent their money on posters, public events, TV and radio spots, direct mailing as well as TV and media appearances.

The “hot phase” of the campaigns started after a TV debate between Winfried Kretschman and Guido Wolf on January 14, 2016, eight weeks before election day. The hot topics of the campaign, besides the personal competition between Winfried Kretschmann and Guido Wolf, were the European refugee crisis and, related to that, the sexual assault incidents of New Year’s Eve 2016 in Cologne and other big cities in Germany.

Cooperation with the Green Party

I approached the Green party’s campaign manager in Baden-Württemberg and the party’s Secretary General in Berlin in late October 2015, to convince them to implement a field experiment within their campaign. I received a 6,000 Euro budget to experimentally test the effects of direct mail on turnout and Green vote share. It was at my discretion to randomly assign the treatment. I also designed the content of the treatment, which was upon approval by party officials, as the letters were part of the actual campaign communication.

Contacts with the Deutsche Post and a direct marketing agency (SSM Mannheim) and first calculations showed that with the budget of 6,000 Euros and a price per letter of about 0.25 Euros⁸ a maximum of about 24,000 letters could be sent out. Now it was up to me to design an experiment to estimate the effect of direct mail on turnout and vote choice.

A Partisan Valence Treatment

To design a partisan valence treatment I studied the campaign material of the Green party. As the whole campaign was focused on the lead-candidate Winfried Kretschmann and his qualities and successes as incumbent minister-president, it was not too hard to develop a valence treatment from the existing campaign materials. The main keywords in the valence treatment are *success* and *reliability*. The main message of the letter was: voters should vote for the Green party so that the successful and reliable minister-president

⁸Including postage, printing and envelopes.

Winfried Kretschmann can continue his work in the upcoming five years. Furthermore, the letter, addressed to the citizens of Baden-Württemberg, emphasized the successes of the incumbent government in the fields of economy, ecology, education and a diverse society. The original treatment, which was approved without changes, can be found in the Appendix. The envelopes were plain white and did not reveal the sender of the letter. The letters were delivered by regular mail.

Another part of the treatment is the timing⁹ of the delivery. My budget limited my efforts to only one wave of letters. Therefore, I wanted to deliver the treatment as close as possible to election day. Thus, the letters were delivered between Tuesday, March 8 2016 and Friday, March 11 2016, with the election being held on Sunday, March 13 2016.

An Ideal Experiment

After the careful definition of my treatment, let me describe a hypothetical ideal experiment following Imai, King, and Stuart (2008) and its implications for my study of the effects of direct mail. The following steps are necessary:

1. “Selecting a large number of units randomly from a well-defined population of interest” (490).
2. “[M]easuring and blocking on all known confounders X” (490).
3. “[R]andomly assigning values of the treatment variable T” (490).

For my study of the effect of direct mail on turnout and party vote choice in Baden-Württemberg an ideal experiment would look like the following:

1. The well-defined population of interest would be all eligible voters in Baden-Württemberg of which a large number (e.g., 5,000) are randomly selected.
2. Measuring confounders, e.g., past voting behavior, how much mail one receives, age and level of education. The next step would be blocking individuals based on those covariates.

⁹Previous research on campaign effects showed that campaign effects are rather short lived (see Gerber and Green 2001).

3. Finally, one would randomly assign individuals from each block to the treatment and control group, thus, creating independent, completely randomized experiments within each block.

Towards a Matched Pair Randomized Design

Unfortunately, it is hard, if not impossible, to implement an ideal experiment¹⁰. Particularly three constraints were limiting my project: *monetary* constraints, *administrative* constraints and *practical* constraints. Therefore, I have to deviate from the ideal experiment in some regards. I will now describe these decisions in detail.

The first and most important design decision was whether to treat and measure at the individual level or at a cluster level. Deviating from the ideal, I chose to assign and measure the treatment at a clustered level. There are two reasons for this. The first reason is an *administrative* constraint: Neither parties nor researchers have access to the voter registration files. It is hard to get access to the population of all eligible voters. The second reason is a *monetary* constraint: For measurement at the individual level I would have to rely on original survey data with two surveys. One survey is needed ahead of the treatment assignment to measure the blocking confounders. A second survey is needed after the election to measure the outcomes at the individual level. Such surveys are expensive. Instead, on different cluster levels both measures on confounders and outcomes are publicly available.

Treating clusters instead of individuals, however, means sacrificing power of the experiment; as for clustered experiments the power of the experiment depends on the number of clusters and not the number of individuals. The biggest disadvantage, however, is that no inferences on individual behavior are possible. To mitigate these disadvantages it is crucial to maximize the number of treated units in the field experiment. I have to identify the smallest cluster units for which the blocking covariates and outcome measures are available, and to which the treatment can be delivered. The smallest cluster units for

¹⁰Imai, King, and Stuart (2008) acknowledge this fact: “Of course, for numerous logistical reasons, ideal experiments are rarely run in practice, and many other research designs are used, depending on the constraints that are imposed by the research situation” (491).

which all three conditions are satisfied are municipalities. The problem with municipalities is, however, that their variation in size is huge¹¹.

To get as many cluster units as possible in the treatment and control groups, I decided to only treat small municipalities. I restricted my population of municipalities to those with at most 1,500 eligible voters in the 2011 elections¹². This leaves me with a list of 185 municipalities all over Baden-Württemberg. The location of those municipalities within the state of Baden-Württemberg is displayed in Figure 1.

In order to increase the power of the experiment, I block my experimental units on known variables that could explain the outcomes¹³ using pair matching¹⁴ following Moore (2012) and Imai, King, and Nall (2009)¹⁵. I generate a list of matched pairs using an *optimal Greedy algorithm with mahalanobis distances*¹⁶. The random assignment is done within each pair. One of the municipalities in each pair is assigned to the treatment group by a virtual coin toss, while the remaining municipality is assigned to the control group. The treatment probability for each of the 185 communities is, thus, $\frac{1}{2}$.

As my budget does not cover the treatment of all the matched pairs, I sort the list of matched pairs from small distances to large to first include more similar pairs. This

¹¹The smallest municipality has just 100 inhabitants and had 72 eligible voters in 2011. The largest municipality has 585,900 inhabitants and had 367,700 eligible voters in 2011.

¹²If people in small municipalities for some reason behave differently than people in the rest of Baden-Württemberg, this comes at the cost of restricting the explanatory power of my experiment to people in small municipalities. Instead of reporting a Population Average Treatment Effect for the whole state, I thus only report a Sample Average Treatment Effect (see Imai, King, and Nall 2009, 32).

¹³Strong predictors are turnout and vote share in previous elections, as well as the socio-demographic profile of a municipality. I also group the municipalities within electoral districts so that treatment and control units always belong to the same electoral district. This is to rule out that candidate effects interfere with the treatment (as voters have only one vote to give to both their district candidate and the state party). The complete list of covariates and their distributions is in the Appendix.

¹⁴Imai, King, and Nall (2009) suggest for cluster randomized experiments: “pair matching should be used whenever feasible” (30). Pair matching is essentially a special case of blocking with two units in each block. Within each pair one unit is randomly assigned to the treatment group while the other unit in the pair ends up in the control group.

¹⁵To implement the pair matching I use the *R* package *blockTools*.

¹⁶For details of the algorithm see Moore (2012).

**The Location of the 185 small communities
in Baden–Württemberg**

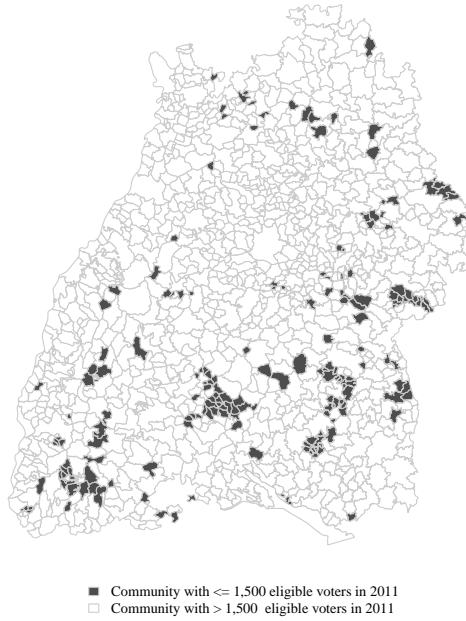


Figure 1. The Location of the 185 small communities in Baden–Württemberg.

procedure leaves me with a sample of 41 municipalities in the treatment group and 41 similar municipalities in the control group¹⁷. The location of the treatment and control municipalities is displayed in Figure 2.

¹⁷The list is in the Appendix.

The Location of the Treatment and Control Communities

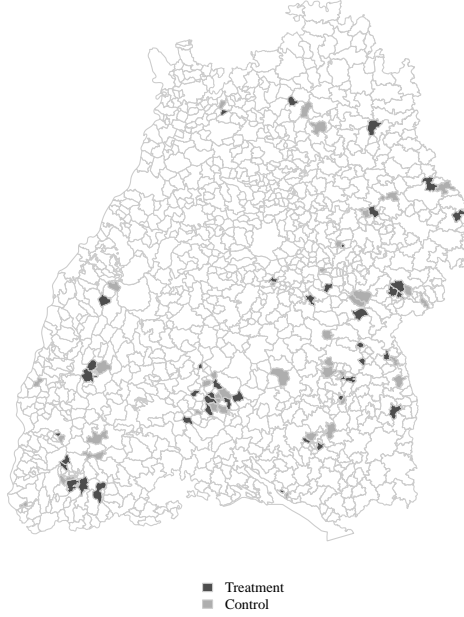


Figure 2. The Location of the Treatment and Control Communities.

Power of the Experiment

To conduct power calculations for matched pair randomized experiments I follow Imai, King, and Nall (2009) who derive power functions for this type of experiment. The expression to calculate the power is:

$$1 - \beta = 1 + \mathcal{T}_{m-1}(-t_{m-1, \frac{\alpha}{2}} | d_U \sqrt{m}) - \mathcal{T}_{m-1}(t_{m-1} | d_U \sqrt{m}), \quad (3)$$

where m is the number of matched pairs,

$\mathcal{T}_{m-1}(\cdot | \zeta)$ is the distribution function of the noncentral t distribution

with $(m - 1)$ degrees of freedom and the noncentrality parameter ζ ,
 $t_{m-1, \frac{\alpha}{2}}$ is the critical value of the one-sample, two-sided level α t-test
 with $(m - 1)$ degrees of freedom,
 and $d_U = \frac{\Psi}{\sqrt{\text{Var}(D_k)}}$,
 where Ψ is the assumed Intent-to-Treat effect and
 D_k the within-pair mean difference in pair k .

With this formula I generate the power curves in Figure 3. To calculate the power levels, I fix α at 0.05 and have to make an assumption about the within-pair mean differences. As an approximation for the post-experiment within-pair mean differences I use the pre-experiment levels of turnout in the matched pairs. The variance of the pre-experiment within-pair mean differences is 0.001. I calculate the power curves for an interval of assumed Intent-to-Treat effects (ITT) starting at 0.1 percentage points up to 1.9 percentage points for every 0.2 percentage points. The number of hypothetical matched-pairs ranges from 2 to 100. The dotted lines in the graph cross at $1 - \beta = 0.8$ and $m = 41$, which is the number of matched pairs I can treat with my budget.

I could thus detect intent to treat effects at the conventional levels of confidence with curves that lie to the upper left of that point. In Figure 3 the minimal detectable effect size lies between 1.3 percentage points and 1.5 percentage points. Precisely, the minimal detectable effect size at a power level of 0.8 and with $m = 41$ matched pairs is at 1.42 percentage points.

Substantively, this means that if my treatment has an effect on turnout and/or the Green vote share I can detect the effect if at least 15¹⁸ per 1,000 eligible voters additionally turnout to vote and/or choose to vote for the Green party in my treatment municipalities as compared to the control municipalities. In total 43,031 people were eligible to vote in 2011 in my treatment municipalities. Thus, in total my treatment needs to generate about 646 additional votes (turnout or Green votes) in order to find an effect. The smallest detectable effect would imply a cost of 9.30 Euros¹⁹ per additional vote, which is a rather high number compared to the total budget of 1,000,000 Euros the Green party spent on their 2016 campaign.

¹⁸Since 14.2 voters per 1,000 voters seems unrealistic.

¹⁹This is the total cost of the experiment divided by the number of additional votes:
 $\frac{6,000}{646}$.

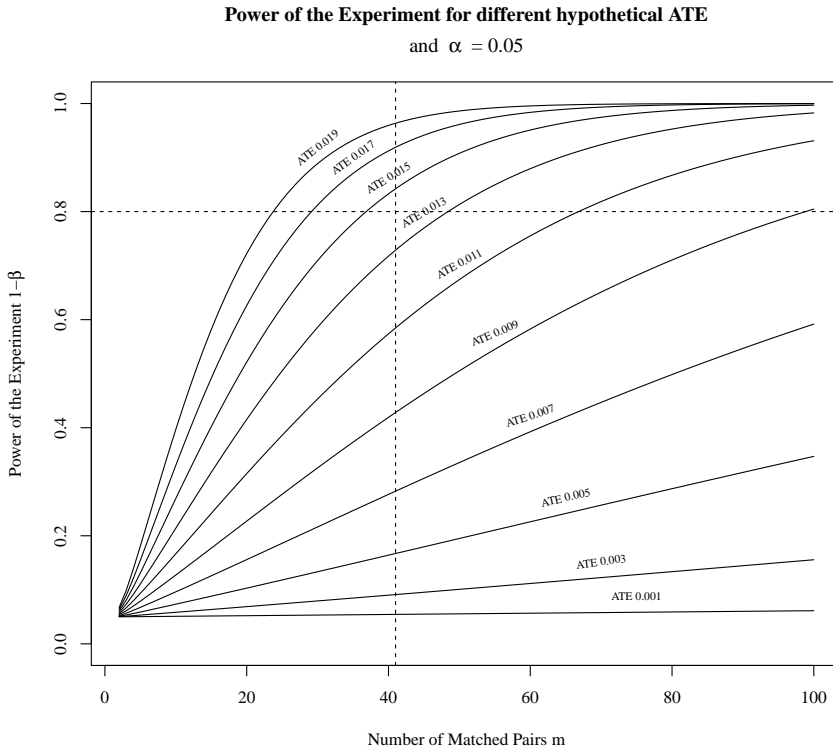


Figure 3. The power of the experiment for 10 different hypothetical ATE and for a different numbers of matched pairs.

Note: the solid lines represent the hypothetical ATE for ten different values. The horizontal dashed line indicates the conventional power level of 0.8. The vertical dashed line indicates the number of matched pairs in my experiment. Thus, the minimal detectable effect lies at the intersection of the two dashed lines.

Election Results

Before turning to the analysis, Table 1 provides a brief overview of the official election results in Baden-Württemberg, as well as the official results in my treatment and control municipalities.

TABLE 1 *The Baden-Württemberg 2016 State Election Results at a glance.*

	Statewide	%	Sample	%	Treatment	%	Control	%
Eligible Voters	7,683,464		83,874		43,093		40,781	
Voters/Turnout	5,411,945	70.4	60,310	71.9	30,848	71.6	29,462	72.2
Valid Votes	5,361,250	99.1	59,655	98.9	30,505	98.9	29,150	98.9
	<i>of which</i>							
Green	1,623,107	30.3	16,126	27.0	8,207	26.9	7,919	27.2
CDU	1,447,462	27.0	21,773	36.5	11,171	36.6	10,602	36.4
AfD	809,564	15.1	8,601	14.4	4,432	14.5	4,169	14.3
SPD	679,727	12.7	5,290	8.9	2,694	8.8	2,596	8.9
FDP	445,498	8.3	4,652	7.8	2,368	7.8	2,284	7.8
other parties	355,892	6.6	3,213	5.4	1,633	5.4	1,580	5.4

RESULTS

Estimating the Intent-to-Treat Effect on Turnout

First, I estimate the effect of the treatment on turnout using a *difference-in-means* approach. For a matched pair design the *difference-in-means* intent-to-treat effect (ITT) estimator is:

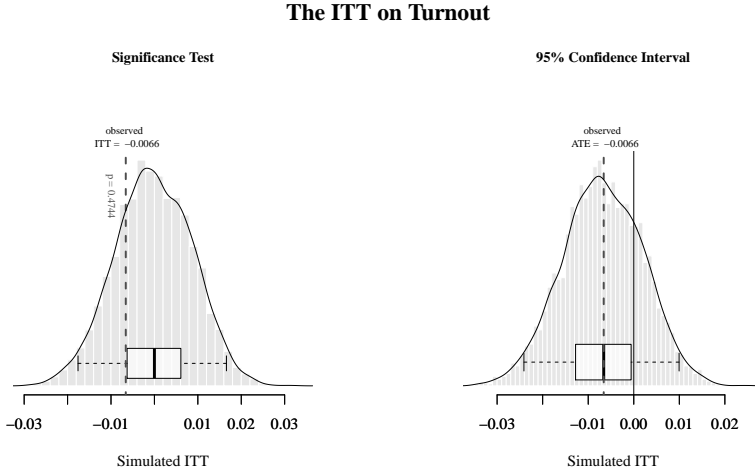
$$\widehat{ITT} = \bar{Y}_{Z=1} - \bar{Y}_{Z=0}. \quad (4)$$

Thus, to obtain the *difference-in-means* ITT on turnout, I subtract the average turnout in my control municipalities (72.24%) from the average turnout in the treatment municipalities (71.56%). This implies a *difference-in-means* ITT of -0.66 percentage points.

To test the sharp null hypothesis I generate a randomization distribution by randomly sampling 5,000 of the 2,199,023,255,552 possible random assignments of the 41 matched pairs. I find that 47.4% of the randomizations under the null yield estimates that are either equal or smaller than -0.66 or equal or larger than 0.66 . The estimated two-tailed p -value, therefore, is 0.474.

In a next step I estimate a 95% confidence interval around the ATE of -0.66 . To this end, I have to use the constant treatment effect assumption to generate the full table of potential outcomes. The confidence interval ranges from -2.4 to 1 . Thus, the *difference-in-means* approach strongly indicates that there is no effect of the treatment on turnout. At the lower end of the confidence interval the treatment decreased turnout by 2.4 percentage points, while at the upper end the treatment increased turnout by 1 percentage

point.



Randomization Distributions from 5,000 simulations of the ITT.

Figure 4. The ITT on Turnout.

Note: the left panel shows 5,000 simulated ITT under the null hypothesis of no effect. The dashed line indicates the observed ITT of the experiment. The whiskers of the boxplot indicate the 2.5 and 97.5 percentile of the null distribution. Thus, it can be seen that the observed ITT is not statistically different from 0. The right panel shows 5,000 simulations under the assumption that the observed treatment effect is the true treatment effect and constant across all experimental units. The whiskers of the boxplot indicate the 95% Confidence Interval for the ITT. The solid vertical line indicates the null hypothesis of no effect and is well within the 95% Confidence Interval.

Figure 4 displays the randomization distributions for the *difference-in-means* significance test, as well as the confidence intervals under the constant treatment effects assumption. In the left panel, the dark dashed line indicates the observed ITT. The whiskers of the boxplot indicate for which ITT the p -value would be exactly 0.05. The right panel, assuming constant treatment effects, shows the confidence interval coverage around the *difference-in-means* ITT estimate. Clearly, the 0 indicated by the black vertical line, is within both bounds of the 95% confidence interval, indicated by the whiskers of the boxplot.

Estimating the Intent-to-Treat Effect on Green Vote Share

I follow the same steps as above using equation 4 to obtain the *difference-in-means* estimate for the ITT on Green vote share. I define Green vote share as: $\text{Green vote share} = \frac{\text{Green votes}}{\text{eligible voters}}$ to get results consistent with the results for turnout.

The actual mean of Green vote share in the treatment municipalities is 19.05% and the mean of Green vote share in the control municipalities is 19.42%. This implies a *difference-in-means* ITT of -0.37 percentage points.

As for estimating the p -value of the ITT on turnout, I generate a randomization distribution by randomly sampling 5,000 of the 2,199,023,255,552 possible random assignments of the 41 matched pairs, to test the sharp null hypothesis of no effect. I find that 46.3% of the randomizations yield estimates that are either equal or smaller than -0.37 or equal or larger than 0.37 . The estimated two-tailed p -value, therefore, is 0.463.

The 95% confidence interval around the *difference-in-means* ATE of -0.37 ranges from -1.3 to 0.6 . This means that the treatment did not affect the Green vote share.

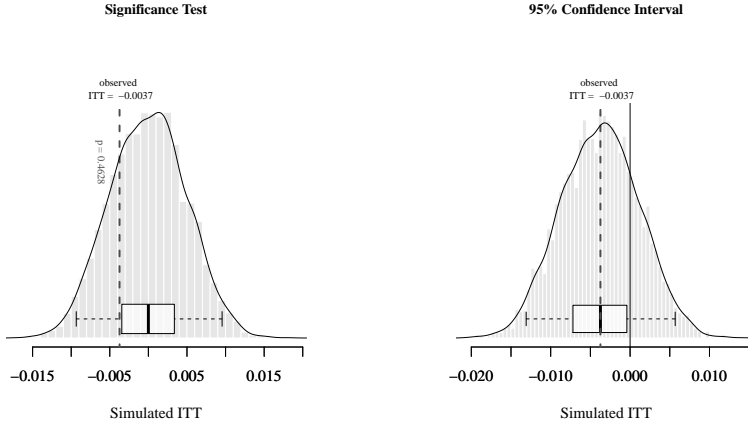
Figure 5 displays the randomization distributions for the *difference-in-means* significance test, as well as the confidence intervals under the constant treatment effects assumption. In the left panel, the dark dashed line indicates the observed ITT. The whiskers of the boxplot indicate for which ITT the p -value would be exactly 0.05. The right panel, assuming constant treatment effects, shows the confidence interval coverage around the *difference-in-means* ITT estimate. Clearly, the 0 indicated by the black vertical line, is within both bounds of the 95% confidence interval, indicated by the whiskers of the boxplot.

What about Non-Compliance?

With any kind of treatment assignment mechanism, treatment compliance is usually imperfect. Angrist, Imbens, and Rubin (1996) define four types of individuals with regard to treatment assignment compliance. In the context of my field experiment, two of their types, *compliers* and *never-takers* are relevant²⁰:

²⁰Irrelevant are *Always-Takers* as they are hard to imagine with the direct mail treatment, at least when I can assume that the Deutsche Post delivered the letters only to treatment municipalities as agreed. Another possibility of *Always-Takers* I can think of are individuals

The ITT on Green Vote Share



Randomization Distributions from 5,000 simulations of the ITT.

Figure 5. The ITT on Green Vote Share.

Note: the left panel shows 5,000 simulated ITT under the null hypothesis of no effect. The dashed line indicates the observed ITT of the experiment. The whiskers of the boxplot indicate the 2.5 and 97.5 percentile of the null distribution. Thus, it can be seen that the observed ITT is not statistically different from 0. The right panel shows 5,000 simulations under the assumption that the observed treatment effect is the true treatment effect and constant across all experimental units. The whiskers of the boxplot indicate the 95% Confidence Interval for the ITT. The solid vertical line indicates the null hypothesis of no effect and is well within the 95% Confidence Interval.

- *Compliers* are the individuals who receive a letter when they live in a municipality that is part of the treatment group and who do not receive a letter when their municipality was assigned to the control group.

who have households in both treatment and control municipalities. However, I assume that this is very few people; *Defiers* are those individuals who would get a letter when their municipality is in the control group and would not get a letter when they live in a treatment municipality. It seems sensible to assume that defiers do not exist in the context of my experiment.

- *Never-Takers* are those individuals who never get the letter, regardless of whether they live in a municipality in the treatment group or a municipality in the control group. I have a considerable share of *Never-Takers* in my municipalities. My direct mail treatment is sent out unpersonalized. Therefore, any individual who opted out from receiving unpersonalized mail does not get the treatment.

So far, in the previous sections I considered only the ITT effects. Another, often reported, quantity-of-interest is the complier average causal effect (CACE)²¹ as introduced by Angrist, Imbens, and Rubin (1996). Basically, the CACE is the ITT estimate adjusted by the share of non-compliers in the treatment group. However, as Schochet and Chiang (2011) note, if one assumes that the same share of non-compliers would have been observed under a different random treatment assignment, the ITT is the “real-world treatment effect” (320). The CACE, thus, is of theoretical value, especially if someone wants to gain knowledge about an individual level mechanism of a treatment.

However, as I have no data on the individual level effects and I assume that the share of non-compliers is about the same in the treatment and control municipalities²², I stick to the real-world treatment effects I estimated above²³.

Checking for Heterogenous Treatment Effects

Lastly, I am interested in whether the treatment effects vary. It is theoretically possible that the treatment had a positive effect in some municipalities, while it had a negative effect in other municipalities. On average I would still find no treatment effect if the positive and negative effects cancel out. Thus, I follow the model-free approach proposed by Ding, Feller, and Miratrix (2016) to test for treatment effect heterogeneity beyond the observed ATE.

Their method builds on the randomization inference approach. However, instead of testing for a *difference-in-means*, they test “whether the treatment outcome distribution

²¹In economics the CACE is known as the local average treatment effect (LATE).

²²Data from the Deutsche Post shows that the share of households who don’t want to receive unpersonalized direct mail is virtually the same (about 15%) in the treatment municipalities and control municipalities.

²³Substantially, it does not make any difference whether I base my interpretation of the results on the ATE or the CACE.

is the same as the control outcome distribution shifted by the average treatment effect” (656). If the distributions differ only in a constant shift, this is a strong indication for constant treatment effects. If the shape of the distributions is different, however, this hints at heterogenous treatment effects. As the true treatment effect is unknown, they propose: “To correct for this, we first construct a confidence interval (CI) for the ATE, repeat the FRT [Fisher Randomization Test] procedure pointwise over that interval and then take the maximum p-value” (656). Theoretically, one could use any test statistic testing for a difference in distributions. Ding, Feller, and Miratrix (2016) prefer a shifted Kolmogorov-Smirnov (SKS) test statistic. With this test statistic a low p-value ≤ 0.05 means that the null hypothesis of similar distributions has to be rejected. Greater p-values indicate that there is no significant difference between the distributions, and therefore no heterogenous treatment effects²⁴. First, I plugged in my data with turnout as the outcome variable. A p-value of 0.3041 indicates that there is probably no treatment effect heterogeneity. Then, I repeated the test with Green vote share as the outcome variable. Again a p-value of 0.4991 shows that heterogenous treatment effects on Green vote share are unlikely.

DISCUSSION AND CONCLUSION

With a partisan field experiment in a high salience election campaign, I am confident that I shed some light on the effects of partisan campaign tools on turnout and vote choice. I conducted a field experiment to test the effects of partisan direct mail in a German state election.

My results show, as expected from the rational choice model of turnout, that my partisan valence treatment has no effect on turnout. However, contrary to the expectation of the rational choice model of vote choice, the treatment also has no effect on the Green vote share. Assuming that the theoretical expectation is still valid, I offer two possible explanations for my empirical findings:

- It might have been that the treatment was too weak. First, it is possible that the treatment was too weak in regard to the mode of delivery. People get many

²⁴Thankfully, Ding, Feller, and Miratrix (2016) implemented their methods in a function library for R. Thus, it was possible to adapt their method.

unpersonalized advertisements by mail every week. Therefore, it is possible that many people did not pay attention to the treatment. Second, it is possible that the treatment was too weak with regard to its message. The whole campaign of the Green party focused on the quality of Winfried Kretschmann. This means that everyone in the state already got the valence treatment independent of my field experiment. However, if this was the case, to detect message effects, field experimentalists would have to send out unexpected messages. Yet, it is unlikely that partisan campaigns would agree to that.

- Another explanation could be that the power of my experiment was not sufficiently large to grasp the small effects that might be there. However, to detect smaller effects we would need bigger experiments in the future. But, finding smaller effects – at least in the context of this study – would only be of theoretical value. It would not change my substantial suggestion to practitioners.

But still, this null result has important implications. For now, I suggest to practitioners not to send out campaign direct mail in the form they currently do it. Instead, it is necessary to evaluate existing campaign tools more critically. Consequently, this field experiment should be seen as the starting point of evaluating more campaign tools with field experiments in high salience elections. Future research, thus, should test other modes of treatment delivery, more messages and try to implement bigger experiments. On the one hand, it could help political scientists to better understand how people come to vote choices. On the other hand, it could help parties who want to spend their campaign money efficiently and win votes. Both political science and campaign practitioners will benefit from it.

REFERENCES

- Angrist, Joshua D, Guido W Imbens, and Donald B Rubin. 1996. "Identification of Causal Effects Using Instrumental Variables." *Journal of the American Statistical Association* 91 (434): 444–455.
- Ashworth, Scott, and Ethan Bueno de Mesquita. 2009. "Elections with platform and valence competition." *Games and Economic Behavior* 67 (1): 191–216. ISSN: 08998256.

- doi:10.1016/j.geb.2008.11.007. <http://dx.doi.org/10.1016/j.geb.2008.11.007>.
- Barton, Jared, Marco Castillo, and Ragan Petrie. 2014. "What persuades voters? A field experiment on political campaigning." *Economic Journal* 124 (574): F293–F326. ISSN: 00130133. doi:10.1111/ecoj.12093.
- Breining, Thomas. 2016. *Landtagswahl Baden-Württemberg. 14 Fakten rund um die Wahl*. <http://www.stuttgarter-zeitung.de/inhalt.landtagswahl-baden-wuerttemberg-14-fakten-rund-um-die-wahl.ee637fcf-fbf9-4998-a253-08475d2a7072.html>.
- Cardy, Emily A. 2005. "An Experimental Field Study of the GOTV and Persuasion Effects of Partisan Direct Mail and Phone Calls." *The Annals of the American Academy of Political and Social Science* 601 (1): 28–40. ISSN: 0002-7162. doi:10.1177/0002716205278051.
- Dale, Allison, and Aaron Strauss. 2009. "Don't Forget to Vote: Text Message Reminders as a Mobilization Tool." *American Journal of Political Science* 53 (4): 787–804.
- Davenport, Tiffany C. 2010. "Public Accountability and Political Participation: Effects of a Face-to-Face Feedback Intervention on Voter Turnout of Public Housing Residents." *Political Behavior* 32 (3): 337–368. ISSN: 01909320. doi:10.1007/s11109-010-9109-x.
- Ding, Peng, Avi Feller, and Luke Miratrix. 2016. "Randomization inference for treatment effect variation." *Journal of the Royal Statistical Society. Series B: Statistical Methodology* 78 (3): 655–671. ISSN: 14679868. doi:10.1111/rssb.12124. arXiv: arXiv:1412.5000v1.
- Druckman, James N, Donald P Green, James H Kuklinski, and Arthur Lupia. 2011. "Experimentation in Political Science." In *Cambridge Handbook of Experimental Political Science*, edited by James N Druckman, Donald P Green, James H Kuklinski, and Arthur Lupia, 3–26. Cambridge: Cambridge University Press.

- Fieldhouse, Edward, David Cutts, Peter John, and Paul Widdop. 2014. "When Context Matters: Assessing Geographical Heterogeneity of Get-Out-The-Vote Treatment Effects Using a Population Based Field Experiment." *Political Behavior* 36 (1): 77–97. ISSN: 01909320. doi:10.1007/s11109-013-9226-4.
- Foos, Florian, and Peter John. 2016. "Parties are No Civic Charities: Voter Contact and the Changing Partisan Composition of the Electorate." *Political Science Research and Methods*: 1–16. ISSN: 2049-8470. doi:10.1017/psrm.2016.48. http://www.journals.cambridge.org/abstract%7B%5C_%7DS049847016000480.
- Foos, Florian, and Eline A. de Rooij. 2017. "The role of partisan cues in voter mobilization campaigns: Evidence from a randomized field experiment." *Electoral Studies* 45 (2017): 63–74. ISSN: 02613794. doi:10.1016/j.electstud.2016.11.010. <http://linkinghub.elsevier.com/retrieve/pii/S0261379416304590>.
- Gelman, Andrew, Gary King, and W. John Boscardin. 1998. "Estimating the Probability of Events That Have Never Occurred: When Is Your Vote Decisive?" *Journal of the American Statistical Association* 93 (441): 1–9. ISSN: 01621459. doi:10.2307/2669597. [http://www.jstor.org/stable/2669597](http://www.jstor.org/stable/2669597%7B%5C_%7D5Cnhttp://www.jstor.org/stable/pdfplus/2669597.pdf?acceptTC=true). <http://www.jstor.org/stable/pdfplus/2669597.pdf?acceptTC=true>.
- Gelman, Andrew, Nate Silver, and Aaron Edlin. 2012. "What is the probability your vote will make a difference?" *Economic Inquiry* 50 (2): 321–326. ISSN: 00952583. doi:10.1111/j.1465-7295.2010.00272.x.
- Gerber, Alan S., James G. Gimpel, Donald P. Green, and Daron R. Shaw. 2011. "How Large and Long-lasting Are the Persuasive Effects of Televised Campaign Ads? Results from a Randomized Field Experiment." *American Political Science Review* 105 (1): 135–150. ISSN: 0003-0554. doi:10.1017/S000305541000047X. http://www.journals.cambridge.org/abstract%7B%5C_%7DS000305541000047X.
- Gerber, Alan S., and Donald P. Green. 2000. "The Effects of Canvassing, Telephone Calls, and Direct Mail on Voter Turnout : A Field Experiment." *American Political Science Review* 94 (3): 653–663.
- . 2001. "Do Phone Calls Increase Voter Turnout ? A Field Experiment." *Public Opinion Quarterly* 65 (1): 75–85.

- Gerber, Alan S, Dean Karlan, and Daniel Bergan. 2009. "Does the media matter? A field experiment measuring the effect of newspapers on voting behavior and political opinions." *American Economic Journal: Applied Economics* 1 (2): 35–52.
- Gerber, Alan, Donald Green, and Christopher Larimer. 2008. "Social Pressure and Voter Turnout: Evidence from a Large-Scale Field Experiment." *The American Political Science Review* 102 (1): 16. ISSN: 0003-0554. doi:10.1017/S000305540808009X. <http://search.proquest.com.ezproxy.royalroads.ca/docview/214437890>.
- Green, Donald P, and Alan S Gerber. 2015. *Get Out the Vote. How to Increase Voter Turnout*. 3rd ed. Washington D.C.: Brookings Institution Press.
- Imai, Kosuke, Gary King, and Clayton Nall. 2009. "The Essential Role of Pair Matching in Cluster-Randomized Experiments, with Application to the Mexican Universal Health Insurance Evaluation." *Statistical Science* 24 (1): 29–53. ISSN: 0883-4237. doi:10.1214/08-STS274. arXiv: 0910.3756.
- Imai, Kosuke, Gary King, and Elizabeth A. Stuart. 2008. "Misunderstandings between experimentalists and observationalists about causal inference." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 171 (2): 481–502. ISSN: 09641998. doi:10.1111/j.1467-985X.2007.00527.x. <http://polmeth.wustl.edu/mediaDetail.php?docId=632>.
- Kendall, Chad, Tommaso Nannicini, and Francesco Trebbi. 2015. "How Do Voters Respond To Information? Evidence From a Randomized." *American Economic Review* 105 (1): 322–353.
- Liegy, Guillaume, Arthur Muller, and Vincent Pons. 2013. *Porte-à-Porte: Reconquérir la démocratie sur le terrain*. Paris: Calmann-Lévy.
- Michelson, Melissa R. 2003. "Getting out the Latino Vote: How Door-to-Door Canvassing Influences Voter Turnout in Rural Central California." *Political Behavior* 25 (3): 247–263.

- Michelson, Melissa R, and David W Nickerson. 2011. "Voter Mobilization." In *Cambridge Handbook of Experimental Political Science*, edited by James N Druckman, Donald P Green, James H Kuklinski, and Arthur Lupia, 228–240. Cambridge: Cambridge University Press.
- Moore, Ryan T. 2012. "Multivariate continuous blocking to improve political science experiments." *Political Analysis* 20 (4): 460–479. ISSN: 10471987. doi:10.1093/pan/mps025.
- Nickerson, David W. 2008. "Is Voting Contagious? Evidence from Two Field Experiments." *American Political Science Review* 102 (1): 49–57. ISSN: 0003-0554. doi:10.1017/S0003055408080039. http://www.journals.cambridge.org/abstract%7B%5C_%7DS0003055408080039.
- Panagopoulos, Costas. 2010. "Affect, Social Pressure and Prosocial Motivation: Field Experimental Evidence of the Mobilizing Effects of Pride, Shame and Publicizing Voting Behavior." *Political Behavior* 32 (3): 369–386. ISSN: 01909320. doi:10.1007/s11109-010-9114-0.
- Pons, Vincent. 2016. "Will a Five-Minute Discussion Change Your Mind? A Countrywide Experiment on Voter Choice in France."
- Riker, William H, and Peter C Ordeshook. 1968. "A Theory of the Calculus of Voting." *American Political Science Review* 62 (1): 25–42.
- Schochet, Peter Z, and Hanley S Chiang. 2011. "Estimation and identification of the complier average causal effect parameter in education RCTs." *Journal of Educational and Behavioral Statistics* 36 (3): 307–345. ISSN: 1076-9986. doi:10.3102/1076998610375837. <http://ovidsp.ovid.com/ovidweb.cgi?T=JS%7B%5C%7DPAGE=reference%7B%5C%7DD=psyc8%7B%5C%7DNEWS=N%7B%5C%7DAN=2011-11904-002>.
- Shaw, Daron R., and James G. Gimpel. 2012. "What if We Randomize the Governor's Schedule? Evidence on Campaign Appearance Effects From a Texas Field Experiment." *Political Communication* 29 (August 2013): 137–159. ISSN: 1058-4609. doi:10.1080/10584609.2012.671231.

- Stokes, Donald E. 1963. "Spatial Models of Party Competition." *American Political Science Review* 57 (2): 368–377.

APPENDIX

A1. Treatment Letter.



Figure A1. Treatment Letter.

*A2. Summary Statistics of all blocking covariates.**Table A2. Summary Statistics of all blocking covariates.*

Variable	N	Mean	St. Dev.	Min	Max
eligible voters in 2011	82	1,020.134	380.181	206	1,498
turnout in 2011	82	0.674	0.060	0.509	0.778
green vote share in 2011	82	0.182	0.062	0.036	0.393
proportion male	82	0.502	0.012	0.469	0.528
proportion german citizens	82	0.960	0.022	0.900	0.995
proportion under 18	82	0.202	0.021	0.143	0.259
proportion between 18 and 65	82	0.623	0.020	0.581	0.667
proportion over 65	82	0.175	0.026	0.124	0.232

A3. Matched Pairs and Treatment Assignment.

20160216_1_Masterarbeit_finalAssignment_v01

rank in district	Treatment	Control	Multivariate Distance	district
1	1 Kirchheim am Ries	Adelmannsfelden	0.74	Aalen
24	1 Berghülen	Nellingen	0.84	Ehingen
56	1 Hofstetten	Mühlenbach	0.97	Lahr
58	1 Hüg-Ehrsberg	Schallbach	1	Loerrach
78	1 Eschach	Ruppertshefen	1.01	Schwaebisch Gmuend
40	1 Todtmoos	Feldberg (Schwarzwald)	1.07	Freiburg I
9	1 Erlenmoos	Wain	1.17	Biberach
59	2 Wieden	Frönd	1.17	Loerrach
55	1 Neidlingen	Ohmden	1.18	Kirchheim
70	1 Schlattdorf	Altenriet	1.23	Nuertingen
2	2 Eilenberg	Södingen	1.24	Aalen
18	1 Au	Horben	1.25	Breisgau
50	1 Grabenstetten	Mehrstetten	1.29	Hechingen-Muensingen
25	2 Griesingen	Lauterach	1.32	Ehingen
72	1 Ebenweiler	Ebersbach-Musbach	1.45	Ravensburg
87	1 Bärenthal	Mahlstetten	1.45	Tutlingen-Donaueschingen
82	1 Schwenningen	Hettingen, Stadt	1.52	Signauingen
88	2 Frittlingen	Deilingen	1.54	Tutlingen-Donaueschingen
10	2 Seekirch	Tiefenbach	1.55	Biberach
68	1 Roigheim	Widdern, Stadt	1.55	Neckarsulm
26	3 Altheim (Alb)	Merklingen	1.56	Ehingen
49	1 Bärenbach	Böttlingen	1.56	Goepfingen
52	1 Langenburg, Stadt	Zweilingen	1.57	Höhenleithe
85	1 Neidenstein	Spechtbach	1.59	Sinsheim
60	3 Altem	Fischingen	1.6	Loerrach
89	3 Talheim	Hausen ob Verena	1.6	Tutlingen-Donaueschingen
4	1 Obernheim	Nusplingen	1.61	Balingen
41	2 Dachsberg (Südschwarzwald)	Breitnau	1.61	Freiburg I
90	4 Bödingen	Königsheim	1.61	Tutlingen-Donaueschingen
97	1 Hüttisheim	Schnurrupfingen	1.65	Ulm
5	2 Dautmergen	Hausen am Tann	1.67	Balingen
27	4 Weidenstetten	Obermarchtal	1.68	Ehingen
91	5 Bubsheim	Egesheim	1.68	Tutlingen-Donaueschingen
28	5 Oberstadion	Rammingen	1.69	Ehingen
29	6 Altheim	Emersingen	1.73	Ehingen
92	6 Dürbheim	Kolbingen	1.74	Tutlingen-Donaueschingen
54	1 Lautenbach	Seebach	1.76	Kehl
16	1 Daisendorf	Stetten	1.78	Bodensee
38	1 Biederbach	Forchheim	1.85	Emmendingen
73	2 Königseggwald	Hofkirch	1.85	Ravensburg
30	7 Neerstetten	Ballerdorf	1.99	Ehingen
11	3 Kanzach	Moosburg	2.01	Biberach
79	2 Täferrot	Obergröningen	2.02	Schwaebisch Gmuend
6	3 Dormettingen	Weilen unter den Rinnen	2.04	Balingen
12	4 Betzenweiler	Oggelshausen	2.05	Biberach
46	1 Muhlhausen im Tale	Aichelberg	2.08	Geislingen
31	8 Unterstadion	Hausen am Bussen	2.11	Ehingen
13	5 Altenweiler	Gutenzell-Hürbel	2.13	Biberach
93	7 Buchheim	Rengulshausen	2.13	Tutlingen-Donaueschingen
74	3 Boms	Fleischwangen	2.21	Ravensburg
94	8 Reichenbach am Heuberg	Gunningen	2.21	Tutlingen-Donaueschingen
19	2 Wittau	Sölden	2.22	Breisgau
32	9 Setzingen	Asselfingen	2.3	Ehingen
61	4 Hassel	Rummingen	2.33	Loerrach
47	2 Schlatt	Gammelshausen	2.34	Geislingen
42	3 Wutach	St. Märgen	2.37	Freiburg I
66	1 Zwingenberg	Neunkirchen	2.42	Neckar-Odenwald
43	4 Bernau im Schwarzwald	Häusern	2.45	Freiburg I
3	3 Tannhausen	Wört	2.49	Aalen
75	4 Riedhausen	Eichstegen	2.61	Ravensburg
14	6 Alleshausen	Allmannsweiler	2.63	Biberach
100	1 Dettighofen	Eggingen	2.67	Waldshut
22	1 Höfen an der Enz	Rohrdorf	2.7	Calw
67	2 Neckarzimmern	Binau	2.83	Neckar-Odenwald
33	10 Breitingen	Nerenstetten	2.88	Ehingen

Figure A3. List of Matched Pairs and Treatment Assignment.