Personalized Treatment Effects:

Evidence from a Randomized Field Experiment

Jason Sclar

QMSS Thesis

Columbia University

May 3, 2012

Contents

List of Tables

1 Introduction

Over the past half-century the nature of the American electoral system has undergone a major change in the way voters come to understand and interpret their political leaders and the ways in which political leaders signal their identity to their constituents. The older system could be loosely characterized as one in which voters took their selection cues from the organized major political parties. Starting in 1956, there was a dealignment of voters from parties (Campbell et al., 1960). As voters' attachments to parties weakened, the driving force in elections has become the personal connection between voter and candidate. This is far stronger than the contemporary voter and party link for determining electoral choices. To the extent that this is the case, the challenge campaigns face is one of forging a personal connection with their constituents. In this thesis I will explore the effectiveness of mailings as an effective means of forging a personal connection between campaign and voter.

Compared to other Western democracies, the United States has lower strength of party-voter linkage which leads to lower average voter turnout (Powell, 1986). Given this, there is accumulated evidence indicates that social connections play a large roll in people deciding whether or not to vote (Putnam, 1966; Putnam, 2001; Rosenstone and Hansen, 2002). Given this, a personal connection is important for increasing voter turnout. Personal connections have a secondary effect on turnout that is mediated through candidate choice (Powell, 1986). Candidates who fail to make personal connections with voters fail to win office (Fenno, 1978). Hence the choice of communication media is absolutely crucial. As a general proposition, voters are always

more inclined to support a candidate with whom they are familiar. Familiarity in this context is not limited strictly to knowing a candidate's platform and beliefs, the threshold of familiarity may be as low as simply knowing the candidate's name (Jacobson, 1987). Historically, members of congress used some combination of approaches to cement their local political base; they tried to cultivate loyalty through the power of their office to benefit their constituents (Mayhew, 1974) and by cultivating an attractive local persona that they present when they return to their home districts (Fenno, 1978).

At the same time as politicians were busy trying to forge personal connections with voters; their methods of campaigning were growing increasingly impersonal. Electronic media does not have the same sort of personal touch that face-to-face campaigning does. With as many as 300,000 people casting ballots in a single Congressional race, it is not feasible for a candidate to speak with all the voters he will need to win office. In recent years, paid televised advertising and earned media, the practice of getting a candidate or campaign inserted into news stories, have become the dominant forms of campaigning. While it may be less effective than other forms of campaigning, paid and earned media have the broadest coverage of voters on a dollar-for-dollar basis. However, with a recent rise of new technology and available data from commercial vendors like Catalist LLC and Voter Vault, campaigns have begun to seek ways to add a personal touch to their materials to help the effectiveness of their messages (Burton and Shea, 2010). They are doing this both by delivering their messages through new venues like SMS messages and also by customizing the content of the messages based on what they know about the person on the receiving

end. These are both attempts to increase the personalization of the campaign messaging by manipulating both the packaging of the message and the content even as the mass-produced nature of modern campaigning requires a more wholesale approach to the challenge, the attempt is to make the final message appear to be personal.

This personalization of the vote occurs at all levels of politics and yet there is still often insufficient efforts to personalize campaigns. In the past three presidential elections, the winning candidate has been the one who best portrayed himself as able to relate to the concerns of the average voter. Lewis-Beck et al. (2008) define six dimensions that are used to evaluate candidates in a two-person race – 1) attitude towards the first candidate, 2) attitude towards the second candidate 3) the parties' platforms on foreign issues 4) the parties' platforms on domestic issues, 5) the parties as managers of government and 6) the social group basis of politics. Of these it is the first two, attitudes that appears to be most crucial in differentiating winners and losers. In 2000 and 2004, the gap between voter attitudes towards George W. Bush versus Al Gore and John Kerry, respectively, provided much of the explanation of the final outcomes in votes. In 2008, voting outcomes tracked extremely closely with which candidate voters preferred when ranking them on a 100-point feeling thermometer (Abramson et al., 2010).

Indeed this tendency is drawn from the lessons of consumer marketing, where there is a general consensus that personalizing a message leads to improved results (Kettle and Häubl, 2011; de Pechpeyrou, 2009; Keller, 1992). The theory behind message personalization is that it creates a stronger bond between the seller and the consumer, which turns the marketer into a more credible source of information. This

message personalization in an age when we are bombarded with information is believed to help prevent the customer from treating the message as spam (Roman and Hernstein, 2004). To the extent that this is correct it is believed that personalization enhances the persuasive power of the advertising materials. That said, there is no general consensus on what it means to personalize a message (Kemp, 2001). For example, Peppers and Rogers (1997) define personalization as delivering a targeted solution to the consumer based on their information. By contrast, Kotler et al. (2001) say that personalization is a process of linking segmentation, targeting and positioning in the marketplace. This creates a serious problem when analyzing the effectiveness of message personalization. Many studies that purport to analyze the effect of personalization are not comparable because they are looking at different versions of personalization.

Vesanan and Raulas (2006) identify two types of variables in the personalization process, objects and operation. Objects include the customer, data and the marketing output. Operations include interactions, processes such as analyzing the data, and delivery. Using these variables, Vesanen (2007) defines five types of personalization that may be used. These are segment marketing, transparent personalization, adaptive personalization, cosmetic personalization and collaborative customization. Segment marketing is matching a customer with a message by some means that is better than mass marketing. Transparent personalization carries this concept a step farther and actually involves adapting the message based on specific information about each person. Adaptive personalization gives customers the option to pick from a suite of choices and is most often seen in online marketing. Cosmetic personalization is the

changing of the package in which the message is delivered. For campaign materials, this would include delivering them in a manner that may not seem mass produced. Finally, collaborative customization is the idea that the producer has direct input from the consumer while producing the product.

On a voter file, segments can be created using geographic or demographic divisions. For a campaign, this would involve sending different messages based on demographic information or some other crude set of divisions made on a voter file, such as sending a slightly different message to voters in each county of a Congressional district, with each message addressing concerns of that particular county. In practice, transparent personalization is used in microtargeting, when campaigns use commercial data to target messages at the individual level (Cornfield, 2007). An example of this would be a campaign sending a pro-gun message to registered voters who have recently applied for a hunting license, i.e. voters who are likely to support gun-ownership, based on related behavior. This is not frequently seen in campaigns, but is seen in e-commerce when an website gives the user a choice between multiple ads on a landing page, before taking them to their desired site. In practice, this sort of personalization can be seen when a campaign develops its message through focus groups, adjusting the message to what voters say they want to hear.

A weakness in the literature of personalization is that there is a shortage of research that tests different forms of personalization in the same way, to create a useful frame of reference. The result is that research that attempts to understand personalization is not often comparable to other work that claims to be doing the same thing, because they are using different methods for examining different modes of personaliza-

tion. Dholakia and Sternthal (1977) examine the credibility of sources in advertising while Kim et al. (2010) examine emotional responses to different advertisements for the same product. Both articles can credibly claim to study personalization, yet the results cannot be compared to understand the phenomenon as a whole. This paper continues this weakness, by only testing one type of personalization. Randomized GOTV experiments present an opportunity to test different versions of personalization using the same procedures. I hope that it will be the first of several tests of different degrees of personalization using the same method, to create comparable results.

In recent years, there has been a large growth in the use of randomized field experiments by political scientists (Druckman et al., 2006). One major use of field experiments has been to understand to try to accurately establish the effect of campaign communication. Through repeated trials, they have found that door-to-door canvassing has a significant effect (Gerber and Green, 2000; Michelson and Nickerson, 2011; Green et al., 2003) and phone calls from trained staff or volunteers may work with the right target audience (Green et al., 2003; Gerber and Green, 2001 Nickerson, 2006) but it is not guaranteed to do so (McNulty, 2005; Cardy, 2005; Gerber and Green, 2005), while robo-calls do almost nothing (Green and Gerber, 2008).

Experiments that use direct mail pieces have shown variance similar to the results of phone experiments. Some experiments have shown mailers to be of very little effect (Wong, 2005) while others, that used extreme social pressure, found a very large effect (Gerber et al., 2008). While the decision to target a given experiment to a demographic group may be exogenous, such as the case when the experiment is funded

by an outside group with a specific interest, many experiments that are specifically targeted to some demographic group, are done to see how that group responds to the experimental message (Wong, 2005, Trivedi, 2005).

Some of these experiments have used various forms of message personalization as defined by Vesnan, albeit without using his terminology. Transparent personalization is most often used in social pressure experiments as a way to exert and increased amount of pressure. This is clearly visible in the experiment Gerber et al. (2008) conducted in Michigan where the "neighbors" treatment shared the voting history of each person's neighbor with them on the post card they received. In this experiment, the content of each post card was crafted at the individual level, not at the much cruder level that would be used in segment marketing. If this treatment were crafted at the district level, the amount of social pressure involved in the experiment would have been almost none, because the people receiving the postcard would not have personally known the people whose vote history they had been sent. This would likely have produced a much smaller effect The strength of the effect came from the use of extreme social pressure based on the fact that every person receiving the treatment knew the majority of people identified on their mailer.

2   Experiment Design

In this paper, I seek to rigorously test the effect of cosmetic personalization on voter turnout. This was done using a large-scale randomized field experiment, conducted in six states during the 2010 election, in which some registered voters receive

the personalized message, while others, in the control group, receive no message. The message was delivered on a postcard that was hand-written by a volunteer. This satisfies the conditions of cosmetic personalization. Instead of being delivered in a familiar mass-produced campaign mailer, the GOTV message was delivered in a custom produced, personalized letter. This ought to attract attention because a hand written note should stand out in a mailbox from the standard array of catalogs and credit card offers that most people receive, even though the contents of the message were not personalized.

The data used for this experiment originated in a national voter file that came from Catalist LLC, a firm specializing in data about registered voters in all states. For each voter, it has their age, race and vote history. This experiment was conducted in six states: Colorado, Florida, Nevada, Ohio, Pennsylvania and Texas.

The experimental population was selected from a specific segment that was of interest to the organization conducting the experiment. A data set of 3,808,637 voters was drawn from the voter files of the six states being studied. This sample consisted of unmarried Caucasian women and both married and unmarried non-Caucasian women. Marital status was predicted using a marriage likelihood model that was developed by the company from whom the list was purchased. All women were between the ages of 18 and 70. None of the women were considered to be likely voters and all were considered likely to support the issues of the group conducting the experiment. Like marital status, both of these characteristics were determined from likelihood models developed by the list vendor. From this total population, a sample of 10,700 women was selected as the treatment group. The random selection was done by state to

insure an even distribution across states. The remaining 3,797,937 women were left as the control group.

The sample is primarily composed of African-American, Caucasian and Hispanic voters. Of the total sample, 3,478,217 fell into one of those three categories, while 330,420 voters were had other racial identities. In the treatment group, 3,035 identified as African-American, 4,090 as Caucasian and 2,457 as Hispanic with 1,118 identifying as some other race. The distribution across states varied widely because the chosen states have fairly different racial compositions. This can be seen in Table 1.

Table 1: Racial Composition of Sample by State

race	CO	FL	NV	OH	PA	TX	Total
African-American	195	557	220	1,201	488	374	3,035
Asian	0	72	0	0	0	0	72
Caucasian	1,024	517	607	917	595	430	4,090
Central Asian	1	1	1	1	3	1	8
East Asian	53	0	35	35	41	31	195
Hispanic	522	476	498	166	199	596	2,457
Jewish	85	0	40	69	71	27	292
Middle Eastern	14	1	14	30	15	18	92
Native American	6	13	6	10	6	1	42
Other	0	62	0	0	0	0	62
Pacific Islander	5	0	25	1	5	2	38
South Asian	10	0	1	22	19	6	58
Uncoded	85	1	53	48	58	14	259
Total	2,000	1,700	1,500	2,500	1,500	1,500	10,700

This study was designed to study the effect of sending a personalized get out the vote message to voters. The treatment was conducted in a single wave using hand-written post cards. Randomly selected and randomly ordered lists of voters in the

12

treatment group were given to volunteer writers, who wrote as many postcards as they were able. Some postcards where never written and mailed. The tracking of the volunteers work was incomplete and therefore insufficient to account for compliance with this task. Additionally, there is no way to confirm of delivery or who actually read the mail. Because of these two points, the experiment can only measure the effect from the intention to treat (ITT).

3   Analysis

This section examines the effect of personalized postcards on turnout based on partner organizations intent to treat them with a postcard. How effective were they as a tactic for increasing turnout? The voting outcomes were measured using voting records from Catalist LLC. The Catalist LLC data are gathered directly from each state's board of elections and are not subject to measurement error. This is a large improvement over using surveys where

Table 2 indicates that from the nationwide sample, 25.28% of those who were selected to receive the treatment voted, compared to 24.87% in the control group. The chi-squared test on this difference returned a value of 0.9808 with an insignificant p-value of 0.322, so the null-hypothesis of no difference cannot be rejected. The difference in turnout between the two groups may be attributable to chance or other factors. However, this difference in means does not account for the problem of different assignment across states. It also fails to control for state-effects, race or previous

vote history. These factors can be considered using regression analysis.

Table 2: Difference in Mean: Chi-Square Test

g10 \treatment	0	1	Total
0	2,853,544	7,995	2,861,539
	75.13	74.72	75.13
1	944,393	2,705	947,098
	24.87	25.28	24.87
Total	3,797,937	10,700	3,808,637
	100.00	100.00	100.00

Pearson chi2(1) = 0.9808 Pr = 0.322

Tables 3-8 are balance tables that show the average values for all of the covariates separated by treatment and control for each state. These tables indicate that while there are significant differences between state in each variable, the differences within states are quite small. This is the result of doing each state's random assignment separately. However, because the differences within states are so small, using dummy variables for state effects in the regression analysis will correct for the differences between states.

Table 3: Summary Statistics: Colorado Mean by Treatment

treatment	g08	g06	g04	p08	p06	p04	afam	white	hispanic	age
0	.5739708	.0913719	.2906095	.0216314	.0047051	.0223271	.0994566	.5149709	.2632529	38.24769
1	.577	.092	.2765	.0265	.0045	.021	.0975	.512	.261	38.3345

Table 4: Summary Statistics: Florida Mean by Treatment

treatment	g08	g06	g04	p08	p06	p04	afam	white	hispanic	age
0	.579406	.0943146	.3270975	.0247905	.0095595	.0353358	.344153	.3048781	.2715227	40.85478
1	.5676471	.0929412	.3164706	.0217647	.0082353	.0358824	.3276471	.3041176	.28	41.16412

Table 5: Summary Statistics: Nevada Mean by Treatment

treatment	g08	g06	g04	p08	p06	p04	afam	white	hispanic	age
0	.6462014	.1292839	.3024233	.0170493	.021814	.0309616	.1486156	.4014458	.3165629	41.30418
1	.6466667	.1253333	.2946667	.0166667	.0253333	.0346667	.1466667	.4046667	.332	41.00467

Table 6: Summary Statistics: Ohio Mean by Treatment

treatment	g08	g06	g04	p08	p06	p04	afam	white	hispanic	age
0	.5417452	.1079392	.3395579	.1020022	.0185432	.0406772	.470153	.3720702	.0665831	35.89217
1	.5452	.1036	.3344	.1064	.0184	.0408	.4804	.3668	.0664	35.806

Table 7: Summary Statistics: Pennsylvania Mean by Treatment

treatment	g08	g06	g04	p08	p06	p04	afam	white	hispanic	age
0	.5653416	.0900196	.2439421	.1631988	.0076749	.0147314	.3254977	.4026712	.1319408	38.02338
1	.574	.0913333	.2546667	.1566667	.006	.0146667	.3253333	.3966667	.1326667	38.21133

Table 8: Summary Statistics: Texas Mean by Treatment

treatment	g08	g06	g04	p08	p06	p04	afam	white	hispanic	age
0	.3862518	.0656776	.4126961	.3355115	.0404825	.0224622	.2429511	.3051163	.3840185	42.63086
1	.386	.0686667	.4266667	.3333333	.0426667	.026	.2493333	.2866667	.3973333	42.502

Considering these other factors shows a small, but statistically significant effect from the treatment. Tables 9 and 10 show a regression analysis of the relationship between receiving a treatment and voting and the marginal effect of treatment, converted into a percentage. Because the treatment was randomly assigned and is orthogonal to every variable except state there is no concern about creating significance via multi-colinearity (Green, 2013). Adding controls will reduce the size of the standard errors, but will not change the effect size.

Table 9: Estimation results : logit

	B	(SE)	t
g10			
treatment	0.041*	(0.024)	1.680
CO	0.458***	(0.006)	79.321
FL	-0.172***	(0.004)	-46.334
NV	0.386***	(0.007)	51.622
OH	-0.377***	(0.005)	-75.741
PA	-0.372***	(0.005)	-77.628
age	0.023***	(0.000)	239.042
g08	1.249***	(0.003)	430.529
g06	0.606***	(0.004)	141.432
g04	0.469***	(0.003)	166.959
p08	1.223***	(0.003)	357.191
p06	0.545***	(0.008)	68.277
p04	0.030***	(0.008)	4.022
afam	0.232***	(0.003)	73.588
hispanic	-0.270***	(0.003)	-78.991
Intercept	-3.254***	(0.005)	-594.420
N	3808637		

* p<0.05, ** p<0.01, *** p<0.001
Significance levels calculated with a 1-tailed test

While some experiments, especially those that use social pressure, have the po-

Table 10: logit: Changes in Probabilities for g10

	min->max	0->1	-+1/2	-+sd/2	MargEfct
treatment	0.0067	0.0067	0.0067	0.0004	0.0067

tential to create a backlash (Mann, 2010), this treatment did not have any factors associated with a backlash. It consisted of a single postcard with a relatively innocuous get out the vote message and since the effect is one-sided, significance can be tested using a one-tailed test. The effect of the treatment is significant at the 5% level. Converting to marginal effects shows that the treatment increased turnout by 0.67% after controlling for race, state and previous vote history.

In addition to just looking for the average treatment effects, the next step in the analysis was to look for heterogeneous effects. It would be unfair to assume that all voters are the same and respond the same way to a treatment (Imai and Strauss, 2011; Feller et al., 2013). The two types of variation that I look for are based on age and vote history. Do older voters respond more strongly to the treatment because they are more likely to read their mail than younger voters? Conversely, is it possible that younger voters reacted more strongly to the treatment because they are less likely to receive hand-written letters and this treatment has more potential to stand out? For vote history, do voters with more or less vote history respond more strongly? Those with more vote history may already be inclined to vote and there are less people who can be convinced to vote. By contrast, those with less history may just not choose to vote ever.

To look for this heterogeneous effect, I created several different interaction variables for treatment. In the first analysis, I did this based on age. The first inter-

action was a continuous interaction of age and treatment. The second interaction dichotomized age, breaking it at the median. The third interaction dichotomized age, breaking it at 50, attempting to find a generational divide or the interaction effect. Table 11 repeats the regression analysis three times, interacting age with treatment in each of the three different ways. However, the interaction variable was insignificant in each of these models, so we cannot reject the null hypotheses. The strength of the treatment effect goes neither up nor down with age.

Table 11: Interaction Effects

	m2 B/(SE)	m3 B/(SE)	m4 B/(SE)
g10			
treatment	0.054	0.038	0.056
	(0.073)	(0.035)	(0.030)
treatment_x_age	-0.000		
	(0.002)		
age	0.023***		
	(0.000)		
treatment_x_median_age		0.008	
		(0.049)	
age_cut_median39		0.564***	
		(0.003)	
treatment_x_age50			-0.044
			(0.052)
age_cut_50			0.559***
			(0.003)
co	0.458***	0.447***	0.424***
	(0.006)	(0.006)	(0.006)
fl	-0.172***	-0.169***	-0.179***
	(0.004)	(0.004)	(0.004)

* $p<0.05$, ** $p<0.01$, *** $p<0.001$

19

Table 11: Interaction Effects

nv	0.386***	0.389***	0.388***
	(0.007)	(0.007)	(0.007)
oh	-0.377***	-0.420***	-0.445***
	(0.005)	(0.005)	(0.005)
pa	-0.372***	-0.389***	-0.409***
	(0.005)	(0.005)	(0.005)
g08	1.249***	1.235***	1.239***
	(0.003)	(0.003)	(0.003)
g06	0.606***	0.624***	0.627***
	(0.004)	(0.004)	(0.004)
g04	0.469***	0.504***	0.526***
	(0.003)	(0.003)	(0.003)
p08	1.223***	1.234***	1.237***
	(0.003)	(0.003)	(0.003)
p06	0.545***	0.563***	0.557***
	(0.008)	(0.008)	(0.008)
p04	0.030***	0.054***	0.051***
	(0.008)	(0.008)	(0.008)
afam	0.232***	0.223***	0.264***
	(0.003)	(0.003)	(0.003)
hispanic	-0.270***	-0.279***	-0.233***
	(0.003)	(0.003)	(0.003)
Intercept	-3.254***	-2.598***	-2.517***
	(0.005)	(0.004)	(0.004)
N	3808637	3808637	3808637

* $p<0.05$, ** $p<0.01$, *** $p<0.001$

The second set of heterogeneous effects are based on vote history. Arceneaux and Nickerson (2009) propose a model wherein the analysis focuses on voters who are of median propensity, instead of either high or low and find substantial effects using this model to re-examine previous work. In this experiment, there were no high propensity voters. All voters were chosen because they were low propensity voters.

Instead of looking for voters at the middle, the voters in this experiment most likely to be affected based on Arseneaux and Nickerson's finding are those who are the most likely to vote. Since I do not have the voter propensity model that was used to select the sample, I use past vote history instead. For this analysis, I created a vote index that is the total number of elections that a person voted in from 2006 through 2008. Every person in the sample received a vote index score of 0, 1 or 2. A logistic regression that Interacts dummy variables for vote index scores of 1 and 2 with treatment yields Table 12.

This model has significant findings, but it also creates a theoretical problem. The interaction of treatment with having a vote index of 1 is large, positive and significant. The interaction of treatment with having a vote index of 2 is larger still, positive and significant. The treatment had a stronger effect on people who voted in both 2006 and 2008 than it did on people who voted in just one of the two elections. This follows from the models that Niven (2001) and Arceneaux and Nickerson (2009) propose. However, the value of the treatment variable is large, negative and significant, in spite of the presence of interaction terms. This indicates that people who did not vote in either 2006 or 2008 and received a treatment were significantly less likely to vote than those in the control group who did not vote in either election. This experiment did not use any social pressure, so it is unlikely to produce backlash of the kind that Mann (2010) considers. There is evidence that this is an aberration that can be corrected using propensity score matching (Imai, 2005). A further step with this data is to consider Imai's model as a way of interpreting the outcome.

Table 12: Vote History Interaction : logit

	m5		
	B	(SE)	t
g10			
treatment==1	-0.424***	(0.057)	-7.482
vote_index==1	0.338***	(0.009)	36.375
vote_index==2	1.016***	(0.011)	96.108
treatment==1 & vote_index==1	0.596***	(0.064)	9.298
treatment==1 & vote_index==2	0.655***	(0.096)	6.797
voterbase_age	0.023***	(0.000)	238.683
co	0.467***	(0.006)	80.657
fl	-0.170***	(0.004)	-45.608
nv	0.390***	(0.007)	52.139
oh	-0.375***	(0.005)	-75.167
pa	-0.369***	(0.005)	-76.972
g08	0.880***	(0.009)	96.315
g04	0.471***	(0.003)	167.571
p08	1.227***	(0.003)	358.028
p06	0.546***	(0.008)	68.220
p04	0.039***	(0.008)	5.143
afam	0.231***	(0.003)	73.348
hispanic	-0.272***	(0.003)	-79.611
Intercept	-3.237***	(0.005)	-589.263
N	3808637		

p<0.05, ** p<0.01, *** p<0.001

4   Discussion

This experiment shows a small, but significant effect from the cosmetically personalized postcards. The true average treatment effect (ATE) is larger than the ITT effect reported here because there is no way to measure the contact rate of mail so we treat it as 100% instead of using intention-to-treat as an instrument for treatment (Gerber and Green, 2000). In this case, the ITT effect is a conservative effort to measure the ATE because it is not it is not possible to know which members of the treatment group actually received the treatment. Thus, the measured ITT effect would likely have been stronger just from being able to remove those people who were never even sent a treatment from the treatment group. This difference between ATE and ITT effect sizes is one factor that contributed to the lift in voting among those who were treated appearing smaller than it may actually have been.

Additionally, those selected to receive the treatment may have been the wrong kind of voters for producing visible effects. Treatment effects are not necessarily homogeneous across voters of different propensity. A lack of previous vote history is a contributing factor to the potential ineffectiveness of a treatment (Niven, 2001; Arceneaux and Nickerson, 2009). The voter likelihood model that was used to select voters into the data set uses prior vote history as a strong predictor of likelihood of voting in the future. This resulted in the women selected into this data set not having a history of voting. It is possible that if this treatment was repeated on a population that had a higher propensity to vote, that approached 50%, and a more extensive history of voting, there might be a noticeably larger effect.

While cosmetic personalization may be an effective way of producing the desired outcome of voting, this effect, and effect size, was consistent also consistent with other experiments that used non-personalized direct mail GOTV treatment. A meta-analysis of seventy-five experiments, conducted across twenty-eight studies finds that non-partisan direct mail has an effect of about 0.5%, but with very little statistical significance, despite the large sample size (Green and Gerber, 2008). The effect size of this experiment is consistent with that meta-analysis, which makes it very difficult to tell if there was any added effect from cosmetic personalization.

The most obvious way to improve this experiment to unpack the GOTV mail effect from the cosmetic personalization effect is to alter the research design. As conducted, the women receiving the treatment did vote at a slightly higher rate than the ones who were left in the control group, but it is impossible to say whether they were responding to receiving a GOTV treatment or if they were responding even more strongly to this particular treatment due to the cosmetic personalization. Future research to extend this work would benefit from having two or more treatment groups. At a bare minimum, a research design that included both hand-written post cards as well as printed post-cards with a standard GOTV message would allow for the measurement of the difference in effect sizes between a cosmetically personalized message and a standard, non-personalized one. This would produce a more meaningful result for understanding the effect of cosmetic personalization than just measuring the difference between a personalized message and the control group.

Furthermore, the above research design could be used to examine other forms of personalization, especially segment marketing and transparent personalization. By

testing multiple forms of message personalization using the same voting field experiment framework, we can begin to construct a basis for understanding the relative strengths of differing forms of personalization against each other.

A final extension of this work is to test different forms of personalization against each other within a single experiment. By using different treatments that use multiple forms of personalization, future work would be even better able to see both which forms of personalization are the most effective and how they interact. For example, if a person wanted to test cosmetic and transparent personalization, they could use a research design with four treatments. The first group would receive a standard, non-personalized GOTV mailer. This treatment group would give the baseline effect of a mailer on turnout for the population and election being studied. The second group would receive a cosmetically personalized post-card with a standard message on it, much like the post-cards used in this experiment. This group would show the effect of cosmetic personalization. The third group would receive a printed mailer that was transparently personalized; this would be likely to include information from the voter file that is unique to that individual. This group would allow for the measurement of the effectiveness of transparent personalization, by comparing them to the first group, as well as the relative strength of transparent personalization versus cosmetic personalization by comparing it to the second group. The final group would receive a treatment that was both cosmetically and transparently personalized, hand written versions of the third treatment. The purpose of this final treatment is to measure the effect of layering multiple types of personalization into a single treatment. Is the combined effect using two types of personalization in a postcard greater than, less

than or equal to the effect from using the two types of personalization individually in different treatments?

The other remaining question from this experiment is how many dollars it cost per vote generated. This experiment cost $.50 per piece mailed, for a total cost of $5,350. Based on an estimated effect of 0.67%, this experiment generated 72 votes at a cost of $74.31 per vote. When you factor in the volunteer time that was involved in generating the postcards, this particular treatment seems to have been very inefficient given that it is unlikely that the cosmetic personalization provided much, if any, effect beyond the standard effect of a GOTV mailer.

Given the opportunity costs of a program of hand-written cards, both in terms of dollars spent and volunteer-hours used, many campaigns have better options for how to spend their resources. For larger and better financed campaigns, television advertising is almost a better option. This type of treatment would be most useful for a campaign for a smaller office that is not well funded. In such a race, more expensive advertising options like television are not feasible and containing costs for a mail program is of paramount importance. In small municipal races and others of comparable size, cosmetically personalizing mail messages may, in fact, be the most efficient use of dollars for a campaign.

References

Abramson, P. R., J. H. Aldrich, and D. W. Rohde (2010). *Change and Continuity in the 2008 Elections*. Washington, DC: Congressional Quarterly Press.

Arceneaux, K. and D. W. Nickerson (2009). Who is mobilized to vote? a re-analysis of 11 field experiments. *American Journal of Political Science 53*(1), 1–16.

Burton, M. J. and D. M. Shea (2010). *Campaign Craft: The Strategies, Tactics and Art of Political Campaign Management* (4 ed.). Santa Barbara: Preager.

Campbell, A., P. E. Converse, W. E. Miller, and D. E. Stokes (1960). *The American Voter*. New York: Wiley.

Cardy, E. A. (2005). An experimental field study of the gotv and persuasion effects of partisan direct mail and phone calls. *Annals of the American Academy of Political and Social Science 601*, pp. 28–40.

Cornfield, M. (2007). Maximizing microtargeting. *Campaigns and Elections 28*(1), pp.54.

de Pechpeyrou, P. (2009). How consumers value online personalization: A longitudinal experiment. *Direct Marketing: An International Journal 3*(1), pp.35–51.

Dholakia, R. R. and B. Sternthal (1977). Highly credible sources: Persuasive facilitators or persuasive liabilities? *Journal of Consumer Research 3*(4), pp. 223–232.

Druckman, J. N., D. P. Green, J. H. Kuklinski, and A. Lupia (2006). The growth and development of experimental research in political science. *The American Political Science Review 100*(4), pp. 627–635.

Feller, A., C. C. Holmes, and C. B. Mann (2013). Beyond toplines: Heterogeneous treatment effects in randomized experiments. Forthcoming.

Fenno, R. F. (1978). *Home Style: House Members in Their Districts*. Boston: Little, Brown.

Gerber, A. and D. Green (2000). The effects of canvassing, telephone calls, and direct mail on voter turnout: A field experiment. *The American Political Science Review 94*(3), 653–663.

Gerber, A. S. and D. P. Green (2001). Do phone calls increase voter turnout?: A field experiment. *The Public Opinion Quarterly 65*(1), pp. 75–85.

Gerber, A. S. and D. P. Green (2005). Do phone calls increase voter turnout? an update. *Annals of the American Academy of Political and Social Science 601*, pp. 142–154.

Gerber, A. S., D. P. Green, and C. W. Larimer (2008). Social pressure and voter turnout: Evidence from a large-scale field experiment. *American Political Science Review 102*(01), 33–48.

Green, D. and A. Gerber (2008). *Get Out the Vote: How to Increase Voter Turnout.* Washington, DC: Brookings Institution Press.

Green, D. P. (2013). Field experiment handbook. Forthcoming.

Green, D. P., A. S. Gerber, and D. W. Nickerson (2003). Getting out the vote in local elections: Results from six door-to-door canvassing experiments. *The Journal of Politics 65*(4), pp. 1083–1096.

Imai, K. (2005). Do get-out-the-vote calls reduce turnout? the importance of statistical methods for field experiments. *The American Political Science Review 99*(2), pp. 283–300.

Imai, K. and A. Strauss (2011). Estimation of heterogeneous treatment effects from randomized experiments, with application to the optimal planning of the get-out-the-vote campaign. *Political Analysis 19*(1), pp. 1–19.

Jacobson, G. C. (1987). *The Politics of Congressional Elections.* Boston: Little, Brown.

Keller, T. K. (1992). Getting personal with donors, members, and clients. *Nonprofit World 10*(5), pp.20–23.

Kemp, T. (2001). Personalization isn't a product. *Internet Week 1*(864), p.1.

Kettle, K. L. and G. Häubl (2011). The signature effect: Signing influences consumption-related behavior by priming self-identity. *Journal of Consumer Research 38*(3), pp. 474–489.

Kim, H., K. Park, and N. Schwarz (2010). Will this trip really be exciting? the role of incidental emotions in product evaluation. *Journal of Consumer Research 36*(6), pp. 983–991.

Kotler, P., G. Armstrong, J. Saunders, and W. Wong (2001). *Principles of Marketing - Third European Edition*. Harlow: Pearson.

Lewis-Beck, M. S., W. G. Jacoby, H. Norpoth, and H. F. Weisberg (2008). *The American Voter Revisited*. Ann Arbor: University of Michigan Press.

Mann, C. B. (2010). Is there backlash to social pressure? a large-scale field experiment on voter mobilization. *Political Behavior 32*(3), pp. 387–407.

Mayhew, D. R. (1974). *Congress: The Electoral Connection*. New Haven: Yale University Press.

McNulty, J. E. (2005). Phone-based gotv: What's on the line? field experiments with varied partisan components, 2002-2003. *Annals of the American Academy of Political and Social Science 601*, pp. 41–65.

Michelson, M. and D. W. Nickerson (2011). Voter mobilization. In J. Druckman, D. P. Green, J. Kuklinski, and S. Lupia (Eds.), *Handbook of Experimental Political Science*, pp. pp. 228–242. Cambridge University Press.

Nickerson, D. W. (2006). Volunteer phone calls can increase turnout: Evidence from eight field experiments. *American Politics Research 34*(3), pp. 271–292.

Niven, D. (2001). The limits of mobilization: Turnout evidence from state house primaries. *Political Behavior 23*(4), pp. 335–350.

Peppers, D. and M. Rogers (1997). *Enterprise One-to-one: Tools for Competing in the Interactive Age*. New York: Doubleday.

Powell, Jr., G. B. (1986). American voter turnout in a comparative perspective. *The American Political Science Review 80*(1), pp. 17–43.

Putnam, R. (2001). *Bowling Alone*. New York: Simon & Schuster.

Putnam, R. D. (1966). Political attitudes and the local community. *The American Political Science Review 60*(3), pp. 640–654.

Roman, E. and S. Hernstein (2004). *Opt-in Marketing*. New York: McGraw-Hill.

Rosenstone, S. and J. Hansen (2002). *Mobilization, Participation, and Democracy in America*. Longman classics in political science. London: Longman.

Trivedi, N. (2005). The effect of identity-based gotv direct mail appeals on the turnout of indian americans. *Annals of the American Academy of Political and Social Science 601*, pp. 115–122.

Vesanan, J. and M. Raulas (2006). Building bridges for personalization - a process model for marketing. *Journal of Interactive Marketing 20*(1), pp. 1–16.

Vesanen, J. (2007). What is personalization? a conceptual framework. *European Journal of Marketing 41*(5), pp.409–418.

Wong, J. S. (2005). Mobilizing asian american voters: A field experiment. *Annals of the American Academy of Political and Social Science 601*, pp. 102–114.