

[DRAFT] Fall 2019 H&M Experiment

Andy Zack

1/14/2020

Summary

In November and December, 2019 our grantees ran Hearts and Minds programs across a handful of legislative districts in five states to educate working and middle class people about economic issues that impact their lives and to connect them to the conversation happening about these issues in state Capitols. These programs were followed by a live-interviewer telephone survey designed to evaluate the effects of these programs. The survey included questions about name-ID of their elected officials, favorability ratings of these same officials, issue support of key issues in each state, recall of the mail portion of each program, and one question about their views on the overall approach of government.

Key Results

- The program caused a 2.1 percentage point increase in unfavorable ratings in districts where electeds did not support our economic policies. There was little movement in districts where the elected supported our policies.
- The program caused a small increase in voters' ability to rate their legislators.
- There was not a measurable effect on voters' issue opinions in most states. However, in Michigan, we were able to increase support for the overtime policy.
- The program also had a significant effect on mail recall, but even in the treatment group, the recall percentage was low. We did not appear to move voters' overall opinion on their states' general approach to governing.
- In terms of modeling persuasion. The TicketSplitter model appears to be somewhat helpful, but a custom model for our programs may outperform it.

Recommendations

First, we should check how these results compare to previous programs where we contacted folks in both supportive and unsupportive districts, but these results strongly suggest focusing our programs in districts where the elected is not supportive of our issues, since we seem to be able to make more of an impact there.

Secondly, I suggest using this new model, either alone or in conjunction with other targeting criteria to determine our audiences for the 2020 hearts and minds programs. I have not yet scored the entire voter file in our targeted districts using this model. Once that happens, we can see how the scores are distributed among likely voters in each of our districts and the demographics of the audiences in each place.

Survey and Experimental Design

Within each state, we held out a control group that also recieved the phone survey in order to compare the survey results. There were 9020 completed surveys. We filtered out surveys where the age or gender of the respondent did not match the voter file to help ensure that we interviewed the person we were trying to contact. After this filter was applied, we ended up with 8071 surveys in the experiment. The following table shows the breakdown of responses by state and treatment group.

The data was weighted by treatment assignment, education, gender, and legislative district, with weights trimmed at 1/3 and 3.

State	Control	Treatment	TOTAL
AZ	928	1,300	2,228
CO	596	832	1,428
FL	628	1,224	1,852
ME	433	423	856
MI	781	926	1,707
TOTAL	3,366	4,705	8,071

Program Details by State

	Arizona	Colorado	Florida	Maine	Michigan
Timing	Nov 19 - Dec 17	Nov 25 - Dec 17	Oct 7 - Dec 18	Oct 8 - Nov 20	Dec 5 - Dec 17
R/D Breakdown	4 R, 2 D	1 R, 3 D	8 R, 0 D	3 R, 1 D	3 R, 3 D
Audience Targeting	High NM + high ticketsplitter	Broad/everyone	From earlier experiment	High NM + high ticketsplitter	High NM + high ticketsplitter
Votes on Record	Yes	Yes	Yes	Yes	No
# of Mail Pieces	2	4	7	4	3
# of Digital	Saturation for 4 weeks	No digital yet	Saturation for 9 weeks	Saturation for 5 weeks	Saturation for 2 weeks
Other Tactics	Phones, texts w/ images	2 tele-townhalls	Phones, texts, spanish radio, spanish mail/digital	Phones, texts, 2 townhalls	1 round calls and texts
Earned Media	No	No	Yes, lots	Yes, lots	No
Budget	\$162,000	\$111,000	\$346,403	\$81,000	\$162,718
Budget/Person	\$2.22	\$2.85	\$2.17	\$3.24	\$3.87

Baseline Results

Results from the control group on several of the key issues give a sense of where our audience stands, absent any contact from the Impact Project.

- 74% cannot rate their State Representative or Senator. This combines both those who say they have never heard of the official and those who have heard, but say they cannot rate. 18% have favorable views, and 8% have unfavorable views of their elected official.
- 43% say that investing in schools and infrastructure and expanding health care would be a better approach for their state government, while 30% would prefer to keeping taxes low for the middle class, and reduce burdensome regulations. The rest responded “both,” “something else,” or “don’t know.”

Mail Recall

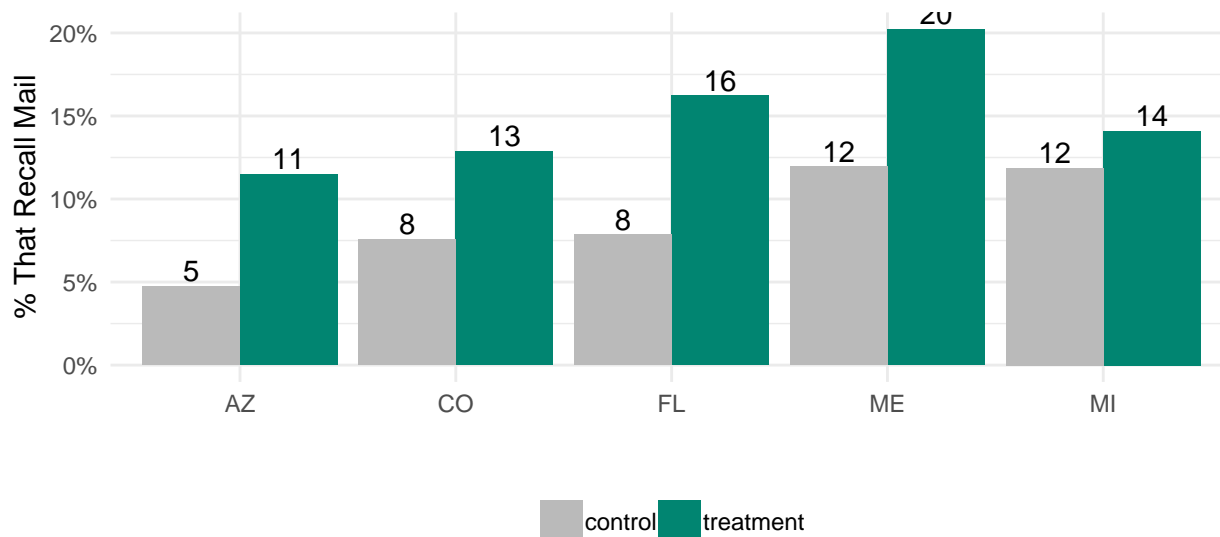
The mail recall question helps put the effects of the program into context.

In the last several weeks, did you receive any mail about issues that the state legislature is working on in [STATE]?

The chart below shows that many people in the treatment group do not recall receiving *any* mail over the several weeks prior to the survey. In all states, the treatment group recalls receiving mail at a significantly higher rate than the treatment group. However, even in Maine, which has the highest rate of recall, only 20% of people in the treatment group remember getting mail. Furthermore, 12% of those in the Maine *control* group say that got mail.

The low rates of recall are not unusual for a program like this. They could be caused by several factors, such as the voterfile address or phone number not matching the correct person, the voter not being the person in

the household who checks mail, survey mismeasurement, or the respondent misremembering. Regardless of the cause, this is a useful metric for putting the rest of the survey results in context.



Favorability

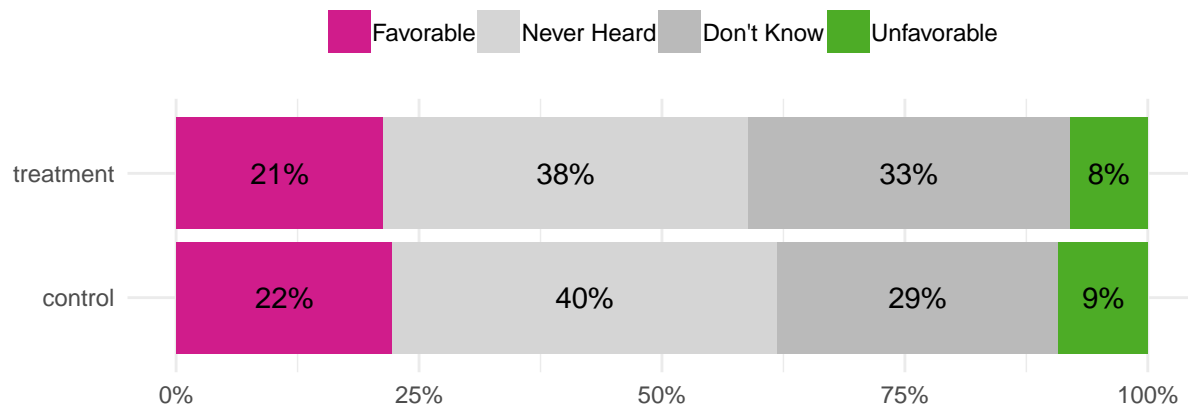
The survey asked respondents to rate the favorability of their elected official as an indicator of whether they truly knew who represented them at the Capitol. Our programs' impact on elected's favorability ratings varies between districts where the elected official supports our issues and districts where the elected does not. For this next section, the analysis is separated into these two groups due to this difference.

The effect on favorability is evaluated in three ways. The first two treat favorability as a *binary* outcome, while the third treats it as a *continuous* outcome.

- Do our programs change the percent of people who give **favorable** ratings?
- Do our programs change the percent of people who give **unfavorable** ratings?
- Do our programs change the **average** favorability, when evaluated on a five-point scale where **strongly unfavorable** = 1 and **strongly favorable** = 5?

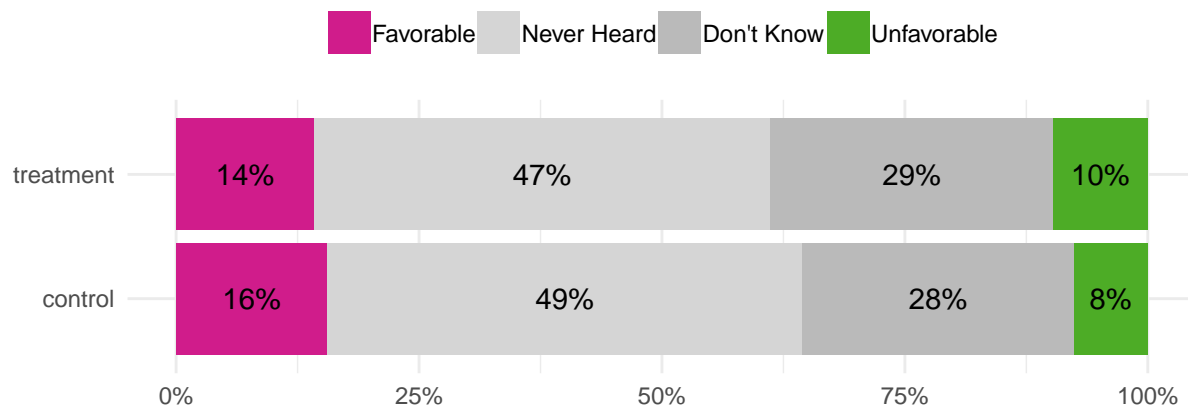
Supportive Districts

In the entire program, we only contacted people living in seven districts that are currently held by elected officials that support our issues. We recieved 2054 responses from people living in these districts, so the estimates here are less precise than those for the other districts. In the Democratic districts, both favorability and unfavorability trended down. However neither of these changes were significant.

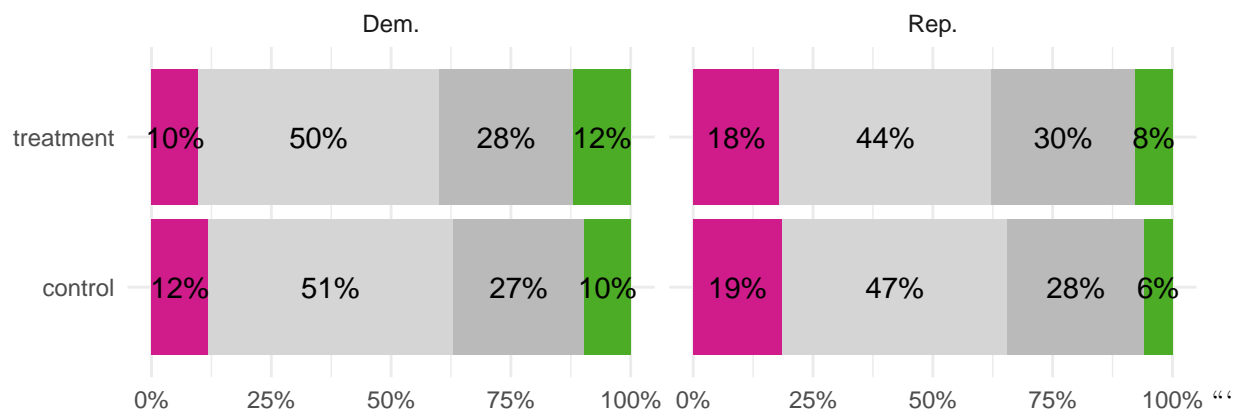


Unsupportive Districts

In the districts where the elected does not support our issues, we did measure significant effects of our program. Unfavorable ratings significantly increased, and the average score on the five-point scale moved towards *strongly unfavorable*.

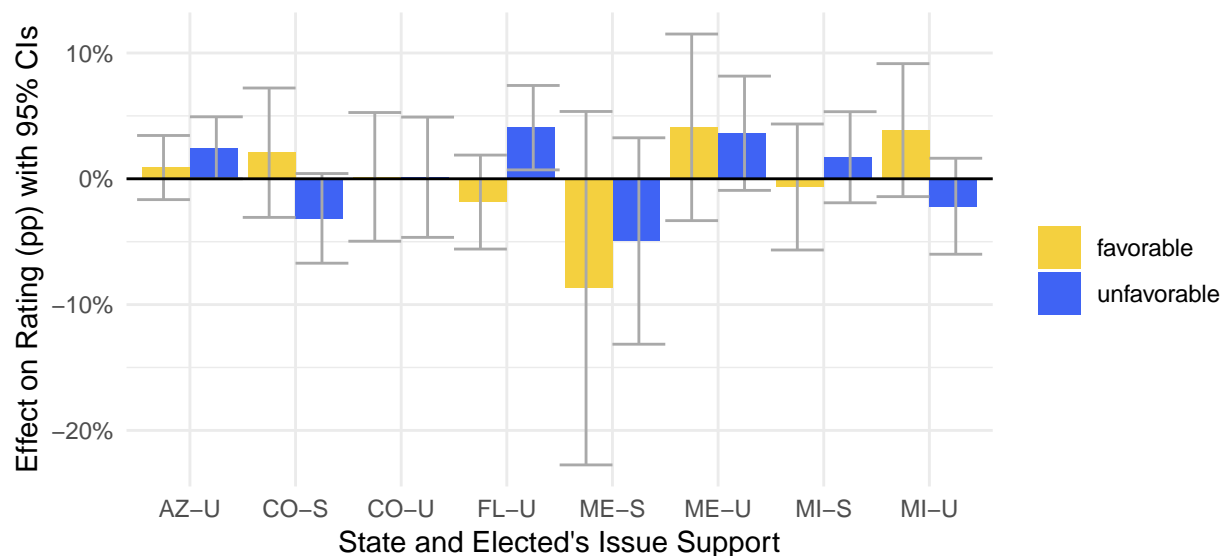


The following table mirrors the plot above (unsupportive districts), but breaks it out by the respondents' party (taken from TargetSmart partisanship scores, rather than actual registration). You can see that movement is fairly similar among voters in both parties.



To further examine our effect on favorable and unfavorable ratings, the following plot shows the effect broken down by state, and further broken out into supportive and unsupportive districts. The effect on unfavorable ratings in Florida and Arizona are both significant for unsupportive districts. The remaining states did not show significant changes. Positive numbers (bars above the 0% line) indicate we increased the number of people given that rating. The grey error bars signify 95% confidence intervals. So, for example, in the

unsupportive districts in Michigan, we increased favorable ratings by 4% and decreased unfavorable ratings by 2%.

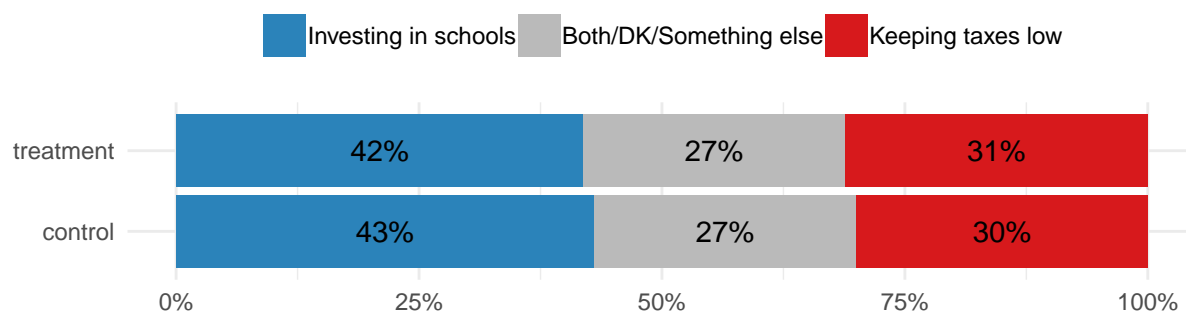


Approach to Governing

Each state's survey included a question to evaluate our program's impact on voter's general approach to governing. Our program caused little change in the answers to this question.

Next, which approach do you think would be better for [STATE]'s government:

- Keeping taxes low for the middle class, and reducing burdensome regulations and taxes, or
- Ensuring big corporations and the wealthy pay their fair share so we can invest in schools, roads, and infrastructure, and expand access to affordable health care



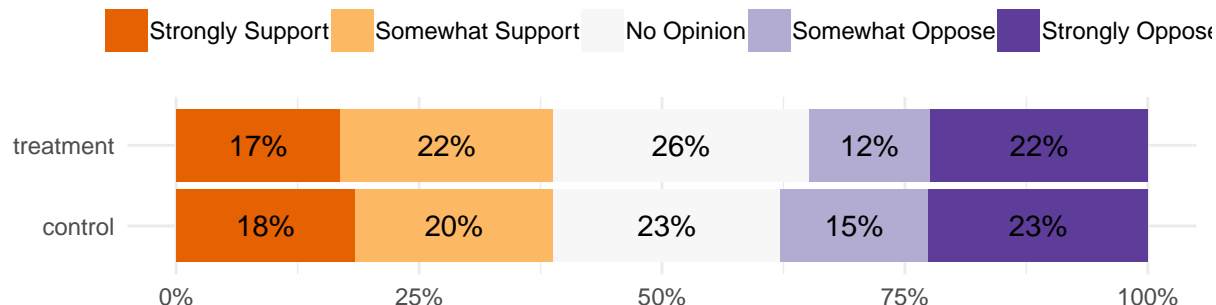
Issues

In each state we asked one or two issue support questions about issues specific to that state. Our programs had very little effect on issue support in nearly every state except for Michigan. There we were able to significantly increase support by 7 percentage points for requiring overtime pay to salaried workers.

Arizona Junk Insurance

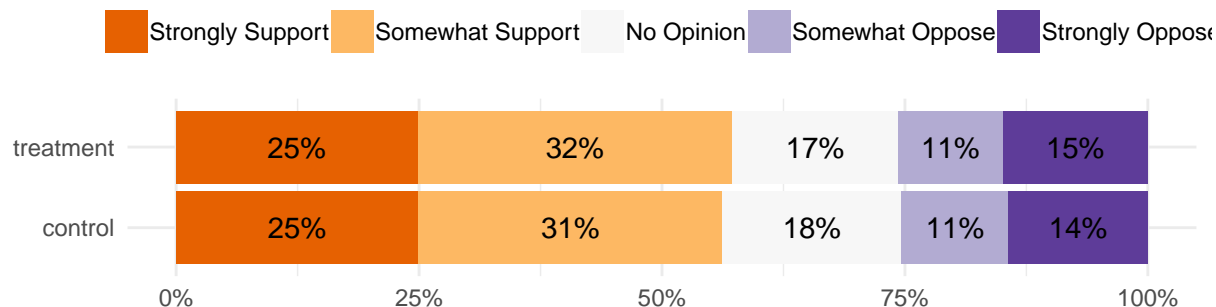
Arizona recently passed a law that allows people to buy inexpensive health insurance plans that offer minimal coverage. Supporters say this reduces health insurance costs, but opponents say

these are “junk” plans and will not offer protections for pre-existing conditions. Do you strongly support, somewhat support, somewhat oppose, or strongly oppose allowing people to buy these low-cost, minimal coverage plans?



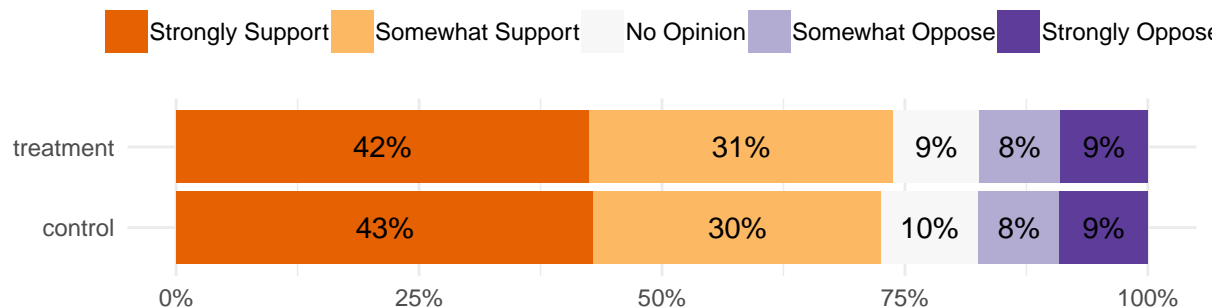
Colorado Retirement

Colorado is considering a law to create a retirement plan that is available to workers who don't have one through their jobs. This would allow employees to save for retirement no matter where they work, but some industry leaders fear it could be a burden on businesses. Do you strongly support, somewhat support, somewhat oppose, or strongly oppose the state creating a retirement plan that is available to all workers?



Colorado Prescription Drugs

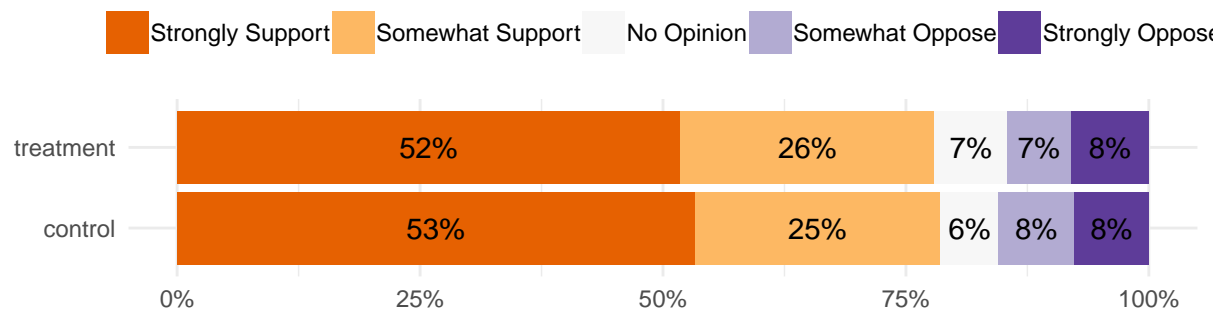
Colorado passed a law that would allow our state to negotiate importing prescription drugs from Canada. This law would make prescription drugs more affordable, but opponents say it could be hard to ensure the safety of such drugs. Do you strongly support, somewhat support, somewhat oppose, or strongly oppose allowing Colorado to negotiate importing drugs from Canada?



Florida Family Leave

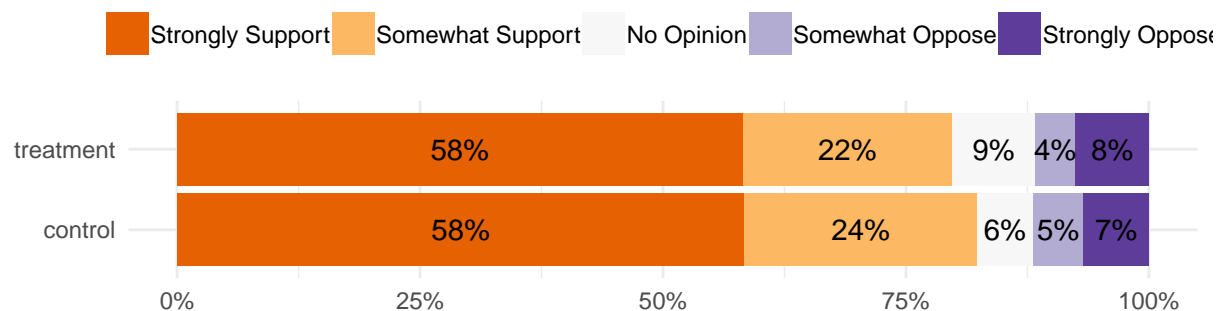
Florida is considering a law to require employers to give their workers paid family leave. This would allow people to take paid time off to care for a new child or sick family member, but

some industry leaders fear it could cost some jobs. Do you strongly support, somewhat support, somewhat oppose, or strongly oppose requiring employers to give employees paid family leave?



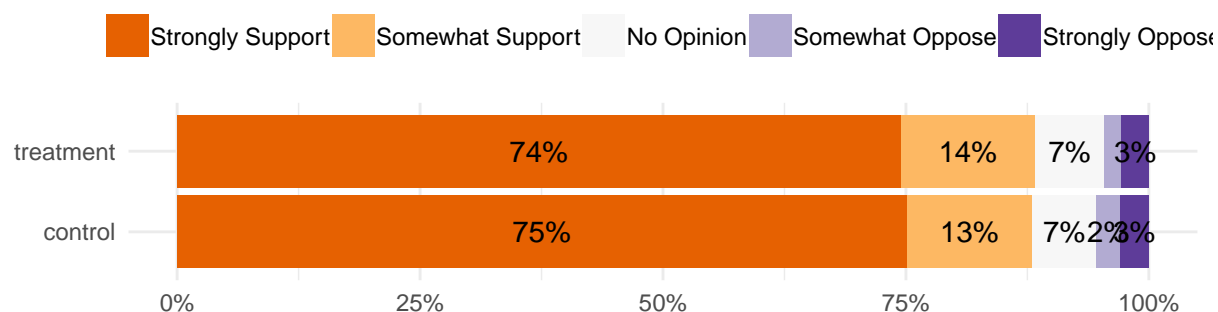
Florida Pre-existing Conditions

Florida is considering another law that would limit how much insurance companies can charge people with pre-existing conditions. This law would make health insurance more affordable for those with pre-existing conditions, but opponents say it could raise costs for others. Do you strongly support, somewhat support, somewhat oppose, or strongly oppose requiring insurance companies to provide more affordable rates for patients with pre-existing conditions?



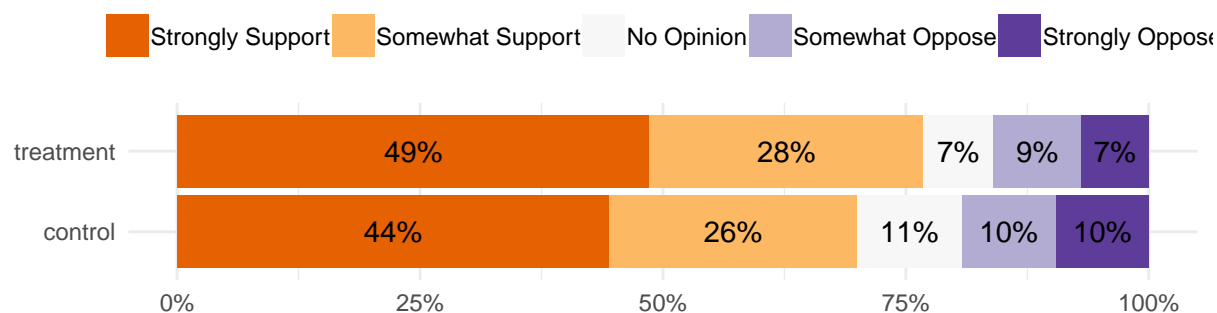
Maine Pay Equity

The Maine state legislature recently passed a law to help ensure that men and women get paid the same amount for equal work. Do you strongly support, somewhat support, somewhat oppose, or strongly oppose this law?



Michigan Overtime

The state of Michigan is considering a rule that would require employers to pay overtime to some salaried workers if they work more than 40 hours in a week. Do you strongly support, somewhat support, somewhat oppose, or strongly oppose this law?



TicketSplitter vs. New Middle Models

In Arizona, Maine, and Michigan, we targeted people who either had a high TicketSplitter score, or a high New Middle score. TicketSplitter is a score created by Catalist that is designed to identify people who are likely to split their ballots. The New Middle scores were created by Benenson and are designed to identify people who care deeply about economic issues, but don't respond to traditional progressive messaging. Both are being used here as proxies for persuadability on our core issues.

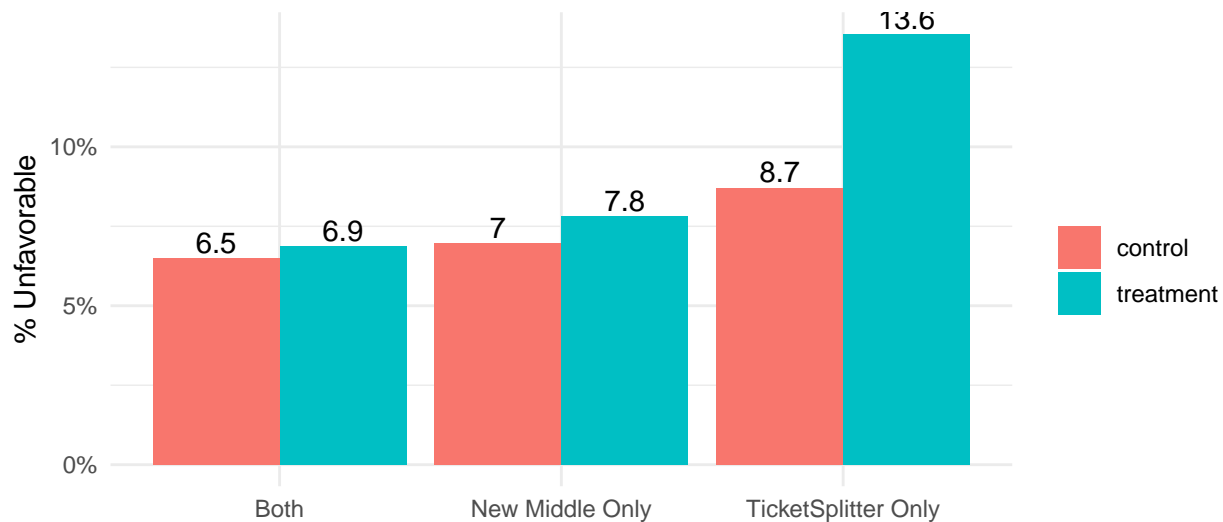
Different New Middle score cutoffs were used in pulling audiences for the different states based on the needs of each state's program. For the purposes of this analysis, I am considering anyone with a score above 33 as New Middle. This roughly corresponds to the 30% of people with the highest New Middle scores.

In these three states plus Colorado where we used very broad targeting, we targeted 2514 people who are in the New Middle Audience, but not the Ticket Splitter audience, 1217 people who are Ticket Splitters, but not New Middle, and 1480 people who are in both audiences.

Ticket Splitter	New Middle	n
No	No	1008
No	Yes	2514
Yes	No	1217
Yes	Yes	1480

The treatment had little effect on unfavorability in the unsupportive districts in these states among people with high New Middle scores. Those with high TicketSplitter scores, but low New middle scores saw a significant increase in unfavorability ratings.

% Unfav. in Unsupportive Districts by Audience and Treatment (AZ, ME, MI)



TIP Persuasion Model

For the final step of this analysis, I built a new model to see how it would compare to the existing models that we had used to target. This model focused only on unsupportive districts, and is built to predict which voters would be more likely to hold unfavorable opinions after receiving treatment from TIP.

Technical Details

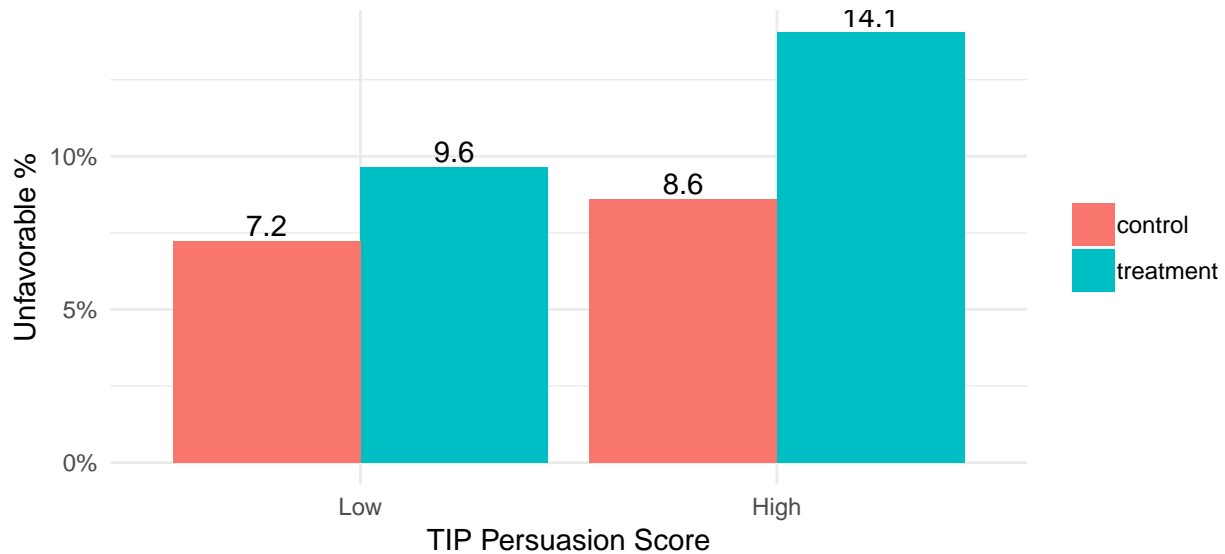
The model was built on a binary outcome where giving an unfavorable rating is the positive response, and giving no rating, or a favorable rating is a negative response. The model was built in R with an X-learner, using a random forest as the base algorithm. This used the causalToolbox package; there is more information on X-learners generally here. The X-learner meta-algorithm is designed to model heterogeneous treatment effects, and it produces conditional average treatment effects (CATEs) which are the estimated average treatment effect within specific subpopulations.

The model used 66 predictor variables. These include basic demographic data from the voter file such as party, race, gender, education, age, and years since registration. It also uses 26 of the TargetSmart scores as inputs; these model support for various issues, as well as turnout, partisanship, and religion. The remaining inputs were various PredictWise scores, household-level variables (such as number of registered voters in the household), and the New Middle Score.

The X-learner was trained using six-fold cross-validation, with CATEs calculated for each fold using the model built on the rest of the dataset. This technique is designed to reduce overfitting.

Model Results

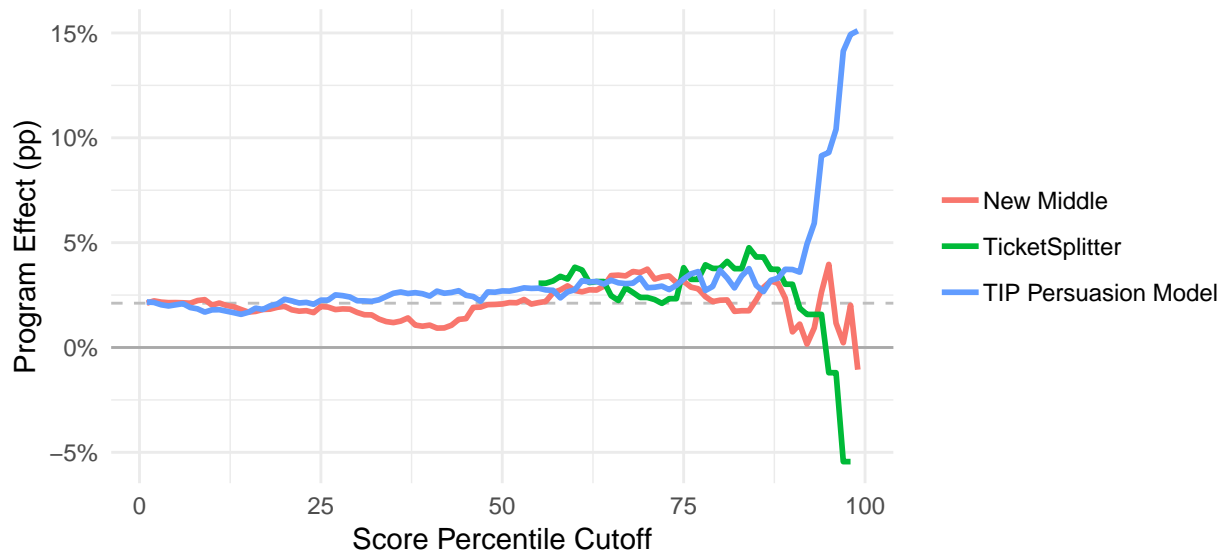
The primary way of evaluating the TIP Persuasion Model is measuring the effect (the difference between the control and treatment group), broken down by people with high and low model scores. The following graph shows this where the 25% of people with the highest TIP persuasion scores are classified as “High,” and everyone else is classified as “Low.” The important part of the graph is comparing the gap between the bars in the low group compared to that gap in the high group. It seems that the program had an effect in both groups, but it was far stronger in those with high TIP persuasion scores. The effect in the high group was 5.5 percentage points, and the effect in the low group was 2.4.



The graph above uses the 75th percentile as a somewhat arbitrary cutoff. Instead, we can examine how effective the program would be for using *any* percentile cutoff. The following graph shows that for the TIP persuasion scores as well as the New Middle Scores and the TicketSplitter scores.

The lines get more unstable, and slightly less accurate towards the right sides of the graph, because in those cases the targeted group has a very small sample size. However there is still a general trend showing the TIP persuasion model outperforming the New Middle model. Ticket Splitter appears to perform somewhat similarly to the TIP persuasion model, but we are missing a lot of data because we only bought this data from Catalyst for people with high TicketSplitter Scores.

The dashed line indicates the overall effect of the program in unsupportive districts. This is what we'd expect the effect to be using the fairly broad targeting that we used in the Fall program. Ideally, any model would increase the effectiveness of the program, so the models should all be above the dashed line, and they should increase towards the right side of the graph, indicating that the programs become more effective as we limit the audience to the best targets. The following graph clearly shows that this is the case for the TIP Persuasion model, but the trend is less clear for the other two.



Note on reading above graph: For example, if we use the 55th percentile as the cutoff, and target the 45% of people with the highest scores, we would have seen an effect of 2.7 percentage points if we used the

TIP persuasion model and an effect of 2.1 percentage points if we had used new middle.

Demographics of TIP Persuasion Model Audience

The following set of tables breaks the voterfile in the targeted districts into three equal sized groups based on scores from the TIP persuasion model. This data comes from the voterfile information that was pulled for this program in November 2019. The numbers may vary slightly with newer information.

The “high” category are the persuadable folks based on this model and they are who we would target if we use this persuasion model.

Age	Low	Mid	High
(17,50]	70%	57%	29%
(50,65]	17%	24%	36%
(65,75]	8%	11%	20%
(75,101]	5%	8%	15%

Gender	Low	Mid	High
Female	45%	49%	57%
Male	51%	47%	41%
Unknown	4%	4%	2%

College	Low	Mid	High
College	37%	33%	32%
Non-College	63%	67%	68%

Partisanship Score	Low	Mid	High
Likely Rep.	55%	54%	40%
Likely Dem.	45%	46%	60%

Party-Registration	Low	Mid	High
Democrat	23%	23%	38%
Other	52%	42%	34%
Republican	26%	34%	28%

Race	Low	Mid	High
Other	7%	8%	11%
Caucasian	86%	76%	64%
Hispanic	7%	16%	25%

Race/Gender	Low	Mid	High
Men of Color	11%	15%	17%
White Men	45%	36%	27%
White Women	38%	37%	37%
Women of Color	7%	12%	20%

Evangelical	Low	Mid	High
No	74%	76%	85%
Yes	0%	0%	3%

Compassion Score	Low	Mid	High
Jerks	27%	34%	33%
Mid	42%	40%	40%
Compassionate	31%	27%	27%

Pres. Turnout	Low	Mid	High
Unlikely Voter	33%	24%	13%
Likely	30%	34%	25%
Very Likely	37%	41%	62%