Online Appendix

A Additional Tables and Figures

Table A1: Included Studies

Number	Article	Issue	Sites	Effect Size[a]	N[b]
1	Broockman and Kalla (2016)	Transphobia	FL	0.207	429
2	Kalla and Broockman (2020, Study 1)	Immigration	CA, TN	0.101	1079
3	Kalla and Broockman (2020, Study 2)	Transphobia	GA, FL, OH, AZ	0.095	1044
4	Kalla and Broockman (2022, Study 1)	Immigration	MI, NC, PA	0.112	1738
5	Kalla and Broockman (2022, Study 2)	Immigration	CA, TN	0.100	529
6	Kalla, Levine and Broockman (2022)	Abortion	ME	0.061	680
7	Santoro and Broockman (2022, Study 1)	Affective Partisan Polarization	Nationwide	0.329	467
8	Santoro and Broockman (2022, Study 2)	Affective Partisan Polarization	Nationwide	0.308	173

[a]Standardized effect size d across all subgroups.
[b]Number of observations ultimately used in meta-analysis.

Table A2: Demographic Characteristics of Study Samples

Study	Issue	Issue	Sites	% White	% Non-White	Mean Age (yrs.)	% Female	% Democrats	% Republicans
1	Broockman and Kalla (2016)	Transphobia	FL	25.349	74.651	49	56.687	48.703	25.349
2	Kalla and Broockman (2020, Study 1)	Immigration	CA, TN	79.240	20.750	52	52.271	34.662	30.955
3	Kalla and Broockman (2020, Study 2)	Transphobia	GA, FL, OH, AZ	88.027	11.973	56	52.203	38.985	28.257
4	(Kalla and Broockman 2022, Study 1)	Immigration	MI, NC, PA	90.017	9.983	53	54.400	29.412	36.034
5	(Kalla and Broockman 2022, Study 2)	Immigration	CA, TN	81.097	18.903	53	49.229	33.360	31.608
6	Kalla, Levine and Broockman (2022)	Abortion	ME	99.433	0.568	48	54.182	14.247	32.826
7	Santoro and Broockman (2022, Study 1)	Affective Partisan Polarization	Nationwide	72.594	27.406	40	50.627	50.0	50.0
8	Santoro and Broockman (2022, Study 2)	Affective Partisan Polarization	Nationwide	78.161	21.839	43	44.828	50.575	49.425

Note: The Table gives measures of partisanship as "% Democrats" and "% Republicans. In all studies except for Study 1 (Broockman and Kalla 2016), these values reflect whether a participant self-identified as a Democrat or Republican on the pre-survey. In Study 1, the pre-survey of which did not ask for partisanship, values for "% Democrats" and "% Republicans" come from voter files.

Table A3: Which Studies Contain Which Subgroups

Subgroup	Studies?
Race	1, 2, 3, 4, 5, 6, 7, 8
Gender	2, 3, 5, 6, 7, 8
Age	2, 5, 6, 7, 8
Education	7, 8
Race-Gender	2, 3, 5, 6, 7, 8
Race-Age	2, 5, 6, 7, 8
Gender-Age	2, 5, 6, 7, 8
Age-Education	7, 8
Gender-Education	7, 8
Race-Education	7, 8
Gender-Age-Education	7, 8
Gender-Race-Education	7, 8
Race-Age-Education	7, 8
Race-Gender-Age	2, 5, 6, 7, 8
Gender-Race-Age-Education	7, 8

Note: Due to data limitations, not every form of demographic concordance ('subgroup') can be studied in every study's dataset. This table lists which studies' data we use to study each form of demographic concordance. (1) is Broockman and Kalla (2016); (2) and (3) are Study 1 and Study 2 in Kalla and Broockman (2020); (4) and (5) are Study 1 and Study 2 in Kalla and Broockman (2022); (6) is Kalla, Levine and Broockman (2022); (7) and (8) are Study 1 and Study 2 in Santoro and Broockman (2022).

Table A4: Meta-regressions in Canvassing Studies

	Race Match		Age Match		Gender Match		Race and Age Match		Gender and Age Match		Race, Gender, and Age Match		Race and Gender Match	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Match	0.00703	-0.0123	0.0223	0.00850	-0.0109	-0.0134	0.0662	0.0558	0.0455	0.0567	0.106	0.0669	-0.0354	-0.0462
	(0.0283)	(0.0282)	(0.0543)	(0.0602)	(0.0296)	(0.0270)	(0.0580)	(0.0492)	(0.0780)	(0.106)	(0.0860)	(0.112)	(0.0467)	(0.0491)
Study 2	-0.0916													
	(0.0793)													
Study 3	-0.108				0.00192								-0.0155	
	(0.0768)				(0.0365)								(0.0499)	
Study 4	-0.0914													
	(0.0765)													
Study 5	-0.109		0.00830		-0.00890		0.00706		-0.0314		-0.0102		-0.0454	
	(0.0802)		(0.0575)		(0.0426)		(0.0515)		(0.0781)		(0.0684)		(0.0600)	
Study 6	-0.119		-0.0272		-0.0564		-0.0497		0.000557		-0.0270		-0.0437	
	(0.0868)		(0.0616)		(0.0462)		(0.0588)		(0.0910)		(0.0773)		(0.0643)	
Study 2 Site 1		-0.0694												
		(0.0866)												
Study 2 Site 2		-0.176		-0.112		-0.104		-0.0954		-0.0300		-0.131		-0.0828
		(0.0821)		(0.0999)		(0.0653)		(0.0625)		(0.194)		(0.168)		(0.0972)
Study 2 Site 3		-0.0185		0.0376		0.00374		0.0241		-0.0574		-0.00139		0.0778
		(0.0851)		(0.0972)		(0.0651)		(0.0645)		(0.190)		(0.152)		(0.0968)
Study 3 Site 1		-0.114				-0.0619								-0.0377
		(0.0783)				(0.0668)								(0.0968)
Study 3 Site 2		-0.0385				0.0335								0.0646
		(0.0850)				(0.0735)								(0.0992)
Study 3 Site 3		-0.142				-0.0821								-0.0528
		(0.0823)				(0.0703)								(0.0967)
Study 3 Site 4		-0.100				-0.0174								0.00177
		(0.0818)				(0.0685)								(0.0969)
Study 4 Site 1		-0.108												
		(0.0770)												
Study 4 Site 2		-0.115												
		(0.0770)												
Study 4 Site 3		-0.0360												
		(0.0779)												
Study 5 Site 1		-0.134		-0.0247		-0.0803		-0.0500		-0.125		-0.0860		-0.0685
		(0.0789)		(0.109)		(0.0681)		(0.0613)		(0.192)		(0.167)		(0.0954)
Study 5 Site 2		-0.0784		-0.0190		-0.0147		0.0286		-0.0902		-0.0795		-0.000537
		(0.0812)		(0.101)		(0.0676)		(0.0640)		(0.192)		(0.153)		(0.0983)
Study 6 Site 1		-0.104		-0.0486		-0.0952		-0.0686		-0.0456		-0.0568		-0.0359
		(0.0780)		(0.0952)		(0.0619)		(0.0620)		(0.188)		(0.144)		(0.0905)
Constant	0.194*	0.198*	0.0981*	0.127	0.105**	0.144*	0.101*	0.125*	0.100	0.149	0.104	0.157	0.126**	0.122
	(0.0731)	(0.0694)	(0.0383)	(0.0768)	(0.0307)	(0.0554)	(0.0341)	(0.0483)	(0.0534)	(0.147)	(0.0445)	(0.113)	(0.0395)	(0.0740)
N Condition-Concordance Pairs	27	27	12	12	20	20	12	12	12	12	10	10	19	19

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: The Table shows the meta-regression results within the canvassing studies. For the fixed effects numbers Study 1 is Broockman and Kalla (2016); Studies 2 and 3 are Study 1 and Study 2 in Kalla and Broockman (2020); Studies 4 and 5 are Study 1 and Study 2 in Kalla and Broockman (2022); Study 6 is Kalla, Levine and Broockman (2022). Study 1 is not listed in the Table because it is the omitted category in the regression when it is present.

Table A5: Meta-regressions in Partisan Animosity Studies

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
	A Match	A and E Match	E Match	G, R, A, and E Match	G, R, and E Match	G Match	G and A Match	G, A, and E Match	G and E Match	G and R Match	G, R, and A Match	R Match	R and A Match	R, A, and E Match	R and E Match
Match ($C = 1$)	0.0721	0.0183	-0.161	0.349	0.0255	-0.210	-0.0626	-0.104	-0.202	0.156	0.184	0.008	0.138	0.184	-0.119
	(0.112)	(0.139)	(0.182)	(0.217)	(0.262)	(0.115)	(0.133)	(0.161)	(0.147)	(0.130)	(0.143)	(0.173)	(0.109)	(0.141)	(0.198)
Study 8	-0.0168	-0.0279	-0.0411	-0.0393	-0.145	-0.041	-0.016	-0.034	-0.088	-0.052	0.023	0.050	0.033	-0.015	-0.051
	(0.152)	(0.155)	(0.200)	(0.149)	(0.247)	(0.152)	(0.155)	(0.149)	(0.161)	(0.157)	(0.145)	(0.182)	(0.129)	(0.137)	(0.204)
Constant	0.293	0.360	0.435	0.339	0.403	0.484	0.363	0.363	0.405	0.314	0.317	0.345	0.301	0.334	0.406
	(0.083)	(0.066)	(0.139)	(0.0629)	(0.174)	(0.083)	(0.069)	(0.063)	(0.081)	(0.073)	(0.066)	(0.141)	(0.075)	(0.065)	(0.149)
N Condition-Concordance Pairs	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

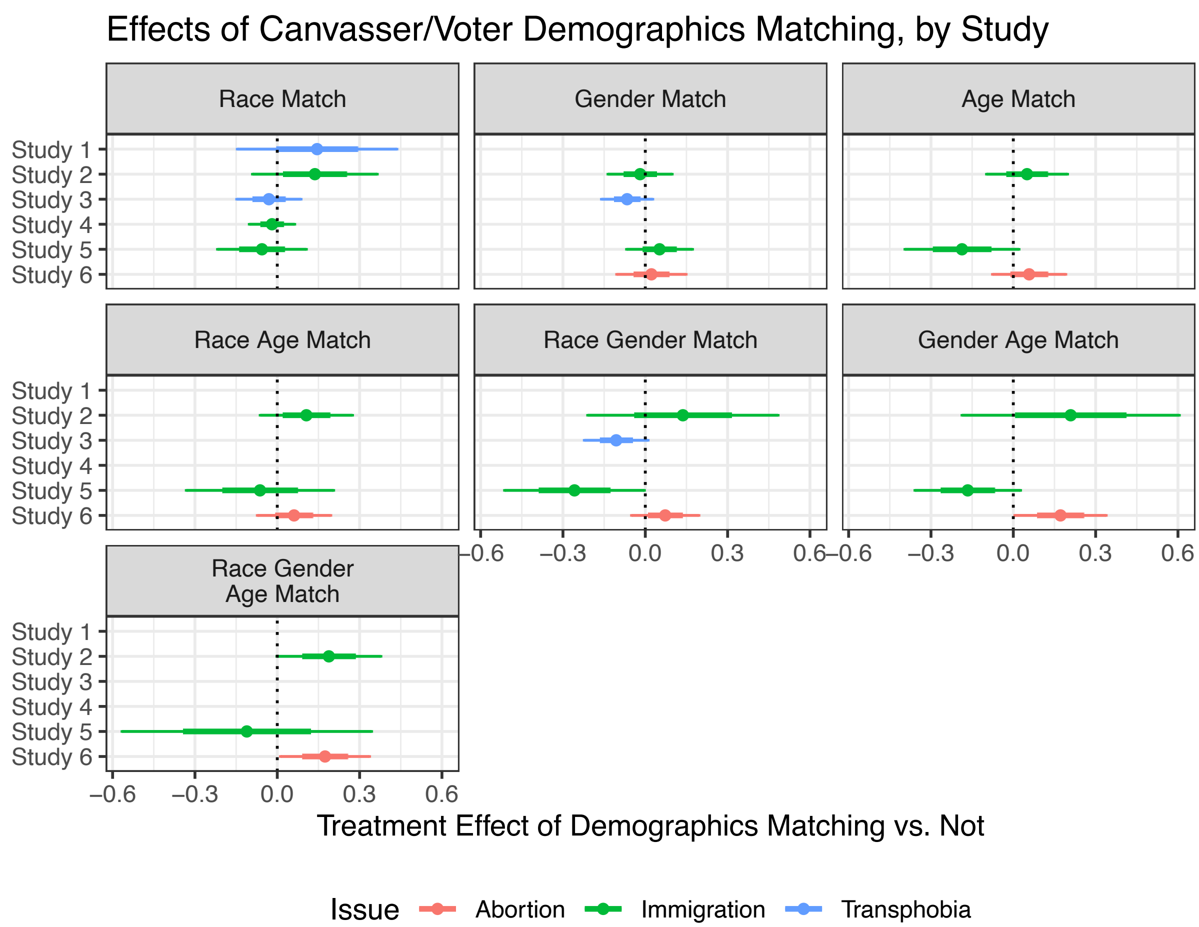
Notes: The Table shows the meta-regression results within the partisan animosity studies: Studies 7 and 8, which are Study 1 and Study 2 in Santoro and Broockman (2022). Study 7 is omitted in the table above because it is the omitted category in the regressions. In the above, A = Age, E = Education, G = Gender, and R = Race.

Table A6: Does demographic matching affect whether a target opens the door?

Demographic Concordance	Estimate	Std. Error
Race	0.020	0.012
Age	0.035	0.056
Gender	0.021	0.029
Race-Gender	0.026	0.039
Race-Age	0.079	0.052
Gender-Age	-0.039	0.041
Race-Gender-Age	0.072	0.064

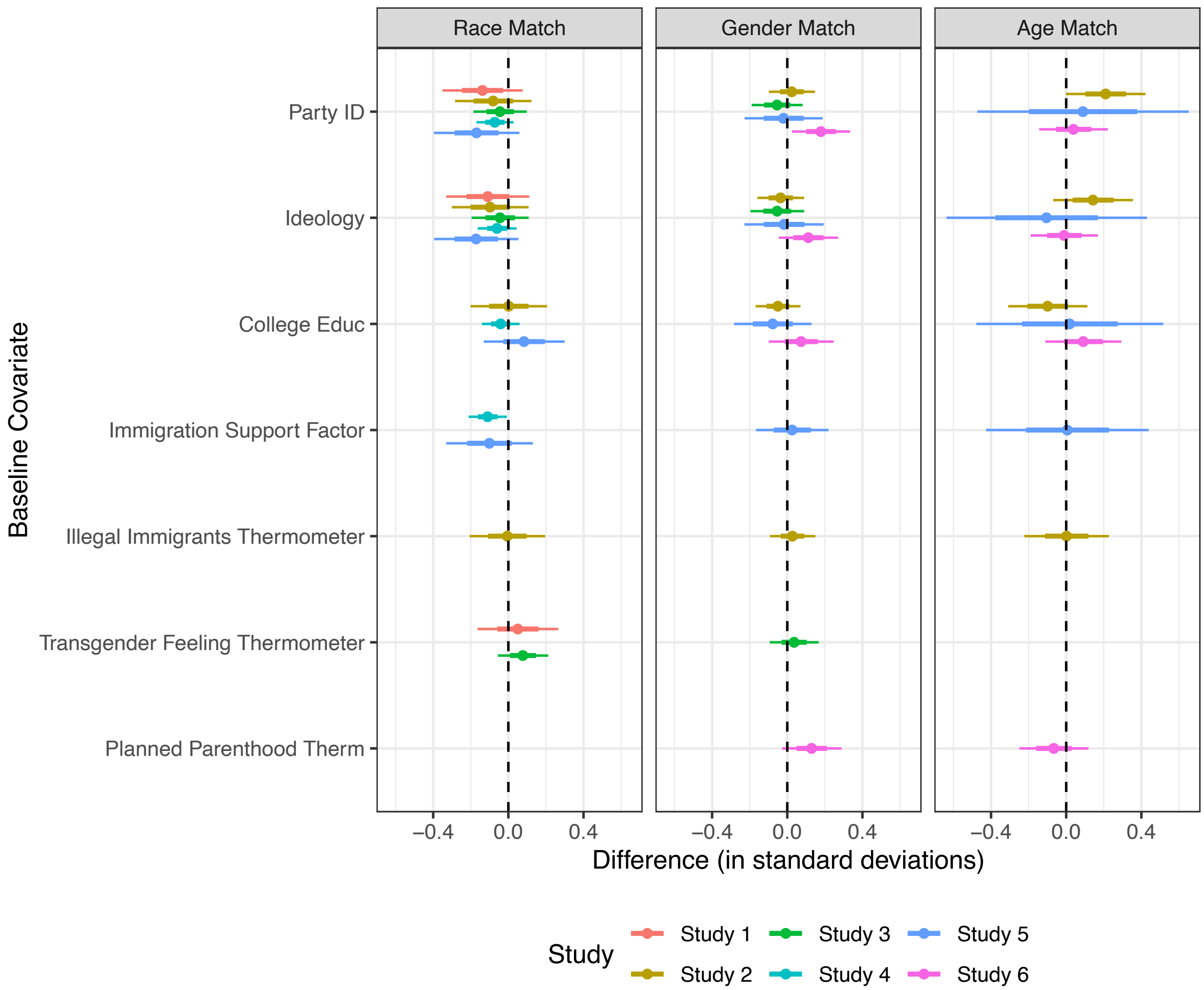
Note: This table tests whether canvassers who did not match their targets demographically were less likely to successfully contact those targets. Data comes from Study 4 (all sites) and Study 5 (TN only), the only studies and sites where data on unsuccessful contacts were available. Each row is the estimate from a separate regression of a binary indicator for whether the door was opened on an indicator for the form of concordance listed in the first column and controls for voters' own race, gender, and age, with experiment-site fixed effects. The coefficient and standard errors on the demographic concordance variables are shown in the table. Standard errors are clustered at the household level. None of the estimates are statistically significant.

Figure A1: Estimates by Study



Notes: The Figure shows the meta-analytic estimates by Study. Point estimates are surrounded by standard errors (thick) and 95% confident intervals (thin). Note that estimates for some combinations of criteria exist even when estimates for the constituent criteria do not when there are insufficient numbers of people who do not satisfy one of the constituent criteria but there are a sufficient number who do not satisfy both; e.g., in Study 6, there are insufficient numbers of cases when race does not match, but there are sufficient numbers of cases when race and age do not both match.

Figure A2: Covariate Balance Between Participants Canvassed by Demographically Concordant and Disconcordant Canvassers in Canvassing Studies



Notes: In data for each of the canvassing studies, linear regressions were run which regressed the covariates shown on the vertical axis on an indicator for the form of demographic concordance shown at the top of each panel and controls for voter race, gender, and age (since demographic concordance's definition is based on voter race, gender, and/or age). Each point above shows the results of one regression. Point estimates are surrounded by standard errors (thick) and 95% confident intervals (thin). 4.4% of the p-values are statistically significant, similar to the 5% of p-values that we would expect to be statistically significant by chance.

B Literature Search and Examples of Excluded Studies

To ensure our meta-analysis was not missing any studies, we conducted a systematic search on Clarivate Web of Science. The Web of Science search was for the terms "persuasion canvassing experiment" and "field experiment persuasion." This resulted in 281 articles. We then examined the individual articles to see whether they fit our inclusion criteria. This search did not yield any studies missing from our analysis; every study did not satisfy at least one of our inclusion criteria.

There are a number of studies that are similar in some ways to the studies we include but lack critical features necessary for our analysis. For example:

- Rossiter (2023). This study uses a text-based chat platform where participants have "real-time written conversations online." Because participants cannot observe the perceived demographics of their interlocutor, this study does not fall within our scope conditions.

- Broockman, Kalla and Sekhon (2017). Although, like others in our meta-analysis, this study concerns political canvassing, the replication package does not contain demographic data for the canvassers and the study authors indicated that they do not have these records. This precludes us from assessing the effects of demographic concordance. (This is also a null result, so it is unlikely to affect our conclusions.)

- West (2023). Respondents were assigned to a treatment involving Zoom interactions or a survey-based control with no face-to-face interactions. There is therefore no control group of face-to-face interactions within which we can compute the presence or absence of demographic concordance. (By contrast, the canvassing studies we include above have variation in demographic concordance in the control group because the control group is still asked to come to the door. This allows us to compare the treatment and control groups conditional on the presence or absence of demographic concordance. However, in this study there is no variation in demographic concordance in the control group, so it is unclear to which groups we would compare the treated groups with and without demographic concordance in order to separately estimate the treatment effects by concordance.)

- Baron et al. (2021). This study involves random assignment to workshops involving, at minimum, 14 participants on each campus. In such a context it is unclear how to compute the presence or absence of demographic concordance, since many groups had a mix of concordant and non-concordant members (e.g., if one person out of a group of 14 were Black, it is unclear how to code such a group).

- Levendusky and Stecula (2021). Like the study above, this study involves conversations in groups of up to seven people. This makes coding the presence or absence of demographic concordance impossible.

- Bailey, Hopkins and Rogers (2016). This study involves random assignment of a persuasion canvassing treatment to registered voters. However, the study's replication data do not include canvasser demographics—meaning that we would not be able to measure concordance—and it is also our understanding that canvassers were not randomized to which voters to canvass.

B.1 Discussion of the File Drawer Problem

It is possible that there are unpublished studies with data on demographic concordance that we could not analyze because they are not published. This could produce bias if the studies that are published and we analyzed have systematically different results than the studies which are not published, a well-known problem called the file drawer problem.

However, the file drawer problem is less likely to bias our results in this case than usual. In particular, the file drawer problem is typically expected to bias meta-analyses because publication of the study (and therefore its availability to be meta-analyzed) is conditional on positive results; if only studies with positive results are published, then a meta-analysis of the results of only the published studies would over-estimate the effects found across all conducted (including unpublished) studies.

This issue should be of less concern here, because the statistic we are interested in is not the main effect of the study, but a heterogeneous effect by canvasser demographic concordance, and because our main finding is a null effect. Given this, the particular form of file drawer bias that would lead us to erroneously estimate a null effect even if there were effects in the underlying population of conducted studies is if authors decided *not* to publish entire studies because they *did* find effects of demographic concordance on persuasion. We expect this is unlikely.

C Further Details on Priors Survey

Because the main text reached the BJPS word limit, we present further details about the priors survey here. 63 individuals were invited to participate via email; 31 responded. 19 respondents were academics, 8 respondents were political practitioners, and 4 respondents did not answer this question. After obtaining consent, the survey first reviewed this context. We then asked respondents, "We next want to know your predictions on how treatment effects vary by whether the canvasser and voter share demographics. Remember, you do not need to consider different contact rates across these groups. We want to know, given that the voter came to the door and began the conversation, is one type of canvasser more effective/persuasive than another type?" Respondents were asked about different demographic concordances (e.g., canvasser and voter are of the same race). Respondents could choose from 11 scale points ranging from 100% more effective when there is demographic concordance to no difference to 100% more effective when there is not demographic concordance.

D Ethical Considerations

The authors declare the original human subjects research containing the survey of academics and campaign practitioners was reviewed and deemed exempt by the [REDACTED] Human Subjects Committee. Participants provided informed consent prior to participating in the survey by reading and responding to a consent statement in Qualtrics. There was no deception in this survey. The authors affirm that this article adheres to the APSA's Principles and Guidance on Human Subject Research. The authors declare no ethical issues or conflicts of interest in this research. This research received no external funding. Research documentation and data that support the findings of this study will be made openly available on Dataverse upon publication.

The data we re-analyze in our main analyses is from previously published studies with publicly available data. The authors of these original studies reported receiving IRB approval and obtaining informed consent from study participants.

E Statistical Precision and Power

As a robustness check, we performed a *post hoc* power analysis to contextualize the results of our meta-analysis. Our meta-analysis did not find a significant effect of demographic concordance above and beyond the effect of the canvassing, contra the expectations of many academics and practitioners. In this section, we report the results of a power analysis to determine whether our null findings are sufficiently precise as to be informative.

Our starting point for this analysis was considering what practitioners would consider to be a meaningfully large effect of demographic concordance for organizations conducting door-to-door canvassing—that is, an effect size that we would want to be able to detect. The cost of door-to-door canvassing is high and would become substantially higher if certain canvassers could not canvass certain respondent pools. We first asked how much greater the cost of this demographically concordant canvassing would be, relative to non-concordant canvassing, to tell us how much more effective concordant canvassing would have to be to "break even." Our inspiration for this exercise was Broockman, Kalla and Sekhon (2017), which details cost considerations for door-to-door canvassing in relation to statistical power.

Broockman, Kalla and Sekhon find that a "standard" (i.e., not necessarily demographically concordant) canvassing visit costs approximately \$3.00. Canvassing costs, however, can easily double or triple based on how much canvassers need to be paid and how much planning a given canvassing initiative requires.[5] Assuming that demographically concordant canvassing would be legally feasible, it would create several additional costs above and beyond standard canvassing, which we expect would put the cost of concordant canvassing above even the high end of the usual canvassing cost distribution:

- Travel—increased distance between houses to be canvassed. If a canvasser can only canvass demographically similar individuals, that canvasser will need to skip more doors. They

[5] See also the discussion of targeted canvassing costs in Green and Gerber (2019, p. 47-49), which notes cost estimates for the per-hour salary of paid door knockers ranging from \$12 to \$17.

will need to travel further between households in order to have demographically concordant conversations.

- Recruitment—specific efforts to recruit canvassers to match particular demographics in an area. For example, if an organization has predominantly Hispanic volunteers but is focused on canvassing in a predominately White neighborhood, that organization will likely face greater recruitment costs to identify White canvassers.
- Training—demographically concordant canvassers might require more training and efforts to motivate canvassers, particularly if the organization needs to turn away people who are intrinsically motivated by or excited about the work, if they do not help match the demographics of the area.
- No economies of scale—there is a fixed cost of training that is amortized across canvassing visits, but this amortization cannot scale if some canvassers are utilized for less time just to cover a particular demographic group.

Conservatively, we estimate that ensuring demographic concordance would double the price of a canvassing visit. For concordant canvassing to be worth implementing, the increase in effectiveness over non-concordant canvassing would need to be proportional to the increase in cost. As we note in the paper, the effect size of a standard door-to-door canvassing intervention is usually about 0.1 standard deviations. An organization that wanted to "break even" using concordant canvassing—such that the effect increase would be parallel to the price increase—would want a minimum effect size of 0.2 standard deviations for concordant canvassing—such that the effect size of concordant canvassing relative to non-concordant canvassing is 100%.

To estimate the statistical power available when meta-analyzing the six field experiments that met our scope criteria, we conducted a simulation study of the 0.1 standard deviations effect size for concordance, as well as effect sizes of 0.05 standard deviations, 0.06 standard deviations, 0.075 standard deviations, and 0.2 standard deviations for additional robustness. To do so, we started with the observed covariate-adjusted treatment effect estimates, standard errors, and sample sizes from the existing studies. As noted above, the approximate average treatment effect for canvassing not designed to be concordant is 0.1 standard deviations. Thus, for the purposes of our power analysis simulations, we randomly generated effect sizes for non-concordant canvassing drawn from a normal distribution around 0.10 (with a standard deviation parameter corresponding to the residual heterogeneity estimate $\hat{\tau}$ in the observed CATE distribution, 0.0287 standard deviations). For concordant canvassing, we randomly generated effect sizes from the normal distribution around 0.15 (corresponding to a 50% average treatment effect of concordance, or 0.05 SDs), 0.16 (corresponding to a 60% ATE of concordance, or 0.06 SDs), 0.175 (corresponding to a 75% average treatment effect of concordance, or 0.075 SDs), 0.20 (corresponding to a 100% average treatment effect of concordance, or 0.1 SDs) and 0.30 (corresponding to a 200% average treatment effect of concordance, or 0.2 SDs) with a standard deviation parameter of 0.0287. We then estimated the same meta-regression models that we used in our analyses, on the partially synthetic data. For each type of concordance and effect size, we performed 10,000 simulated iterations.

The results of our power analysis are provided in table A7. First, Table A7a show whether our observed 95% confidence intervals allow us to rule out the respective effect sizes of 0.05, 0.075, 0.1, and 0.2 standard deviations (table A7a). For instance, we are able to rule out effects of 0.05 standard deviations on race and gender concordance, less than half the size we estimated would be required for race and gender concordance to be cost-effective. Second, we report power estimates for each of those effect sizes (table A7b). We are well-powered ($\geq$ 80%) for detecting what we identify as the cost-effective effect size of interest—0.1 standard deviations—for the concordance in race, age, gender, and race-gender.

Table A7: Precision and power across different effect sizes for concordance

(a) Does the 95% confidence interval rule out this effect size of concordance?

Demographic Concordance	0.05 SDs		0.06 SDs		0.075 SDs		0.1 SDs		0.2 SDs	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Race		✓		✓	✓	✓	✓	✓	✓	✓
Age									✓	✓
Gender	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Race-Gender			✓	✓	✓	✓	✓	✓	✓	✓
Race-Age									✓	✓
Gender-Age									✓	
Race-Gender-Age										
Study Fixed Effects?	×		×		×		×		×	
Study-Site Fixed Effects?		×		×		×		×		×

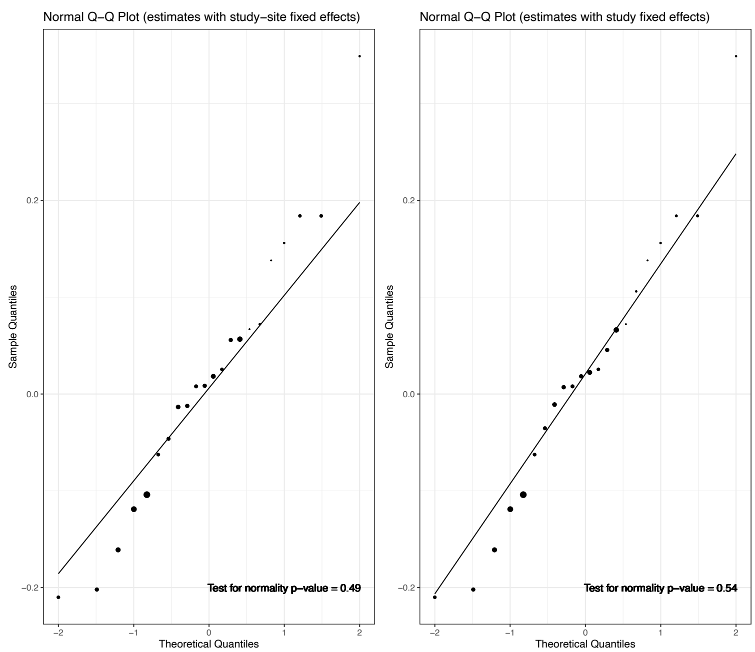
(b) Statistical power for different effect sizes for concordance

Demographic Concordance	0.05 SDs		0.06 SDs		0.075 SDs		0.1 SDs		0.2 SDs	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Race	0.57	0.34	0.83	0.66	0.98	0.95	1.00	1.00	1.00	1.00
Age	0.06	0.05	0.14	0.23	0.40	0.36	0.87	0.84	1.00	1.00
Gender	0.44	0.41	0.74	0.72	0.97	0.96	1.00	1.00	1.00	1.00
Race-Gender	0.12	0.09	0.30	0.22	0.67	0.57	0.98	0.95	1.00	1.00
Race-Age	0.02	0.02	0.05	0.05	0.16	0.16	0.58	0.56	1.00	1.00
Gender-Age	0.00	0.00	0.01	0.01	0.09	0.08	0.51	0.48	1.00	1.00
Race-Gender-Age	0.00	0.00	0.00	0.00	0.03	0.03	0.25	0.25	1.00	1.00
Study Fixed Effects?	×		×		×		×		×	
Study-Site Fixed Effects?		×		×		×		×		×

F False-Discovery Rate Adjustment and Q-Q Plot

Although some point estimates are positive, it is unlikely that these null-but-suggestive results conceal true effects: the distribution of estimates is in line with what we would expect by chance due to sampling variability were all of the true effects null. For instance, when applying a false-discovery rate adjustment Benjamini and Hochberg (1995), *all* of the p-values on our estimates are 1.0. To visualize this, Figure A3 presents a Q-Q plot of all the treatment effect estimates previously shown now plotted against the normal distribution. The distributions match: both visually and using a precision-weighted skewness and kurtosis test of normality, we cannot reject the hypothesis that the distribution of treatment effect estimates is consistent with a normal distribution ($p = 0.49$ with study-site fixed effects; $p = 0.54$ with study fixed effects).

Figure A3: Test for Whether Treatment Effect Estimates are Normally Distributed



G References for Online Appendix

References for Appendices

Bailey, Michael A, Daniel J Hopkins and Todd Rogers. 2016. "Unresponsive and unpersuaded: The unintended consequences of a voter persuasion effort." *Political Behavior* 38(3):713–746.

Baron, Hannah, Robert Blair, Donghyun Danny Choi, Laura Gamboa, Jessica Gottlieb, Amanda Lea Robinson, Steven Rosenzweig, Megan Turnbull and Emily A West. 2021. "Can Americans Depolarize? Assessing the effects of reciprocal group reflection on partisan polarization." Available at https://osf.io/3x7z8/.

Benjamini, Yoav and Yosef Hochberg. 1995. "Controlling the false discovery rate: a practical and powerful approach to multiple testing." *Journal of the Royal statistical society: series B (Methodological)* 57(1):289–300.

Broockman, David E, Joshua L Kalla and Jasjeet S Sekhon. 2017. "The design of field experiments with survey outcomes: A framework for selecting more efficient, robust, and ethical designs." *Political Analysis* 25(4):435–464.

Levendusky, Matthew S and Dominik A Stecula. 2021. *We need to talk: how cross-party dialogue reduces affective polarization*. Cambridge University Press.

Rossiter, Erin. 2023. "The Similar and Distinct Effects of Political and Non-Political Conversation on Affective Polarization." Available at TheSimilarandDistinctEffectsofPoliticalandNon-PoliticalConversationonAffectivePolarization.

West, Emily. 2023. "Cross-Partisan Discussion De-Polarizes White Americans' Racial Policy Preferences." Working paper, available at https://drive.google.com/file/d/1PKH1A3U6jkrJdyok2Kg2H7IRQArKOpvN/view.