Online Appendix for Do Emotional Ads Persuade? Evidence from Real-Time Campaign Advertising Experiments

John Murphy, Alexander Coppock, Seth Hill, and Lynn Vavreck

September 3, 2026

Contents

A Factor analysis: Emotion Indices **2**

B Subgroup Ad Emotional Responses **4**

C Metaregression tables **5**
C.1 Positive emotion index 6
C.2 Negative emotion index 9
C.3 Intended emotion index 12
C.4 Remaking Figure 4 with Target Issue (Stretch Coding) 15
C.5 Remaking Figure 4 with Intended Emotions 16
C.6 Remaking Figure 4 with Target Issue (Stretch Coding) and Intended Emotions . . . 17

D Intended Emotion Codings **18**

E Additional Ratings Analysis **20**
E.1 Assessing pairwise prediction accuracy 21

A Factor analysis: Emotion Indices

In this section, we provide evidence that our negative and positive emotion indices are reliable measures.

Figure A.1 shows a scree plot that includes all four emotion outcomes (happy, hopeful, angry, and worried). The plot provides evidence that these four variables load on two factors that we label positive and negative emotions.

Figure A.1: Scree Plot for Emotion Index Factors

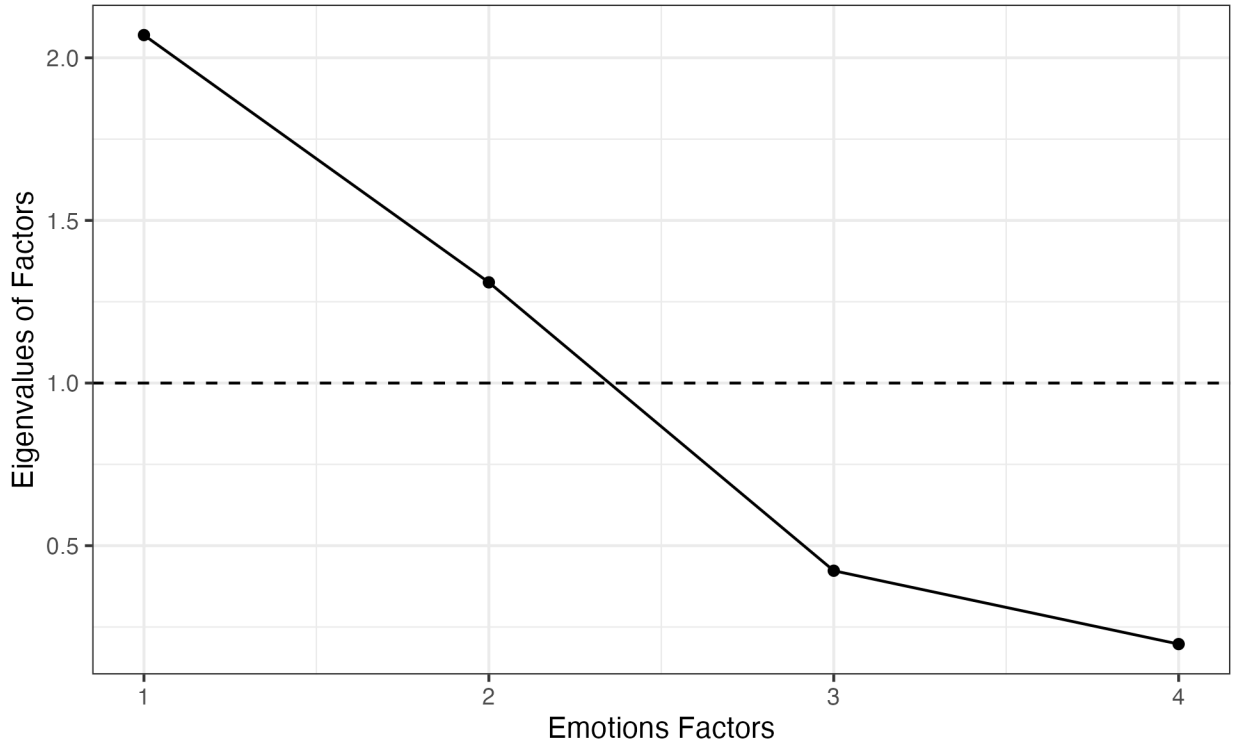
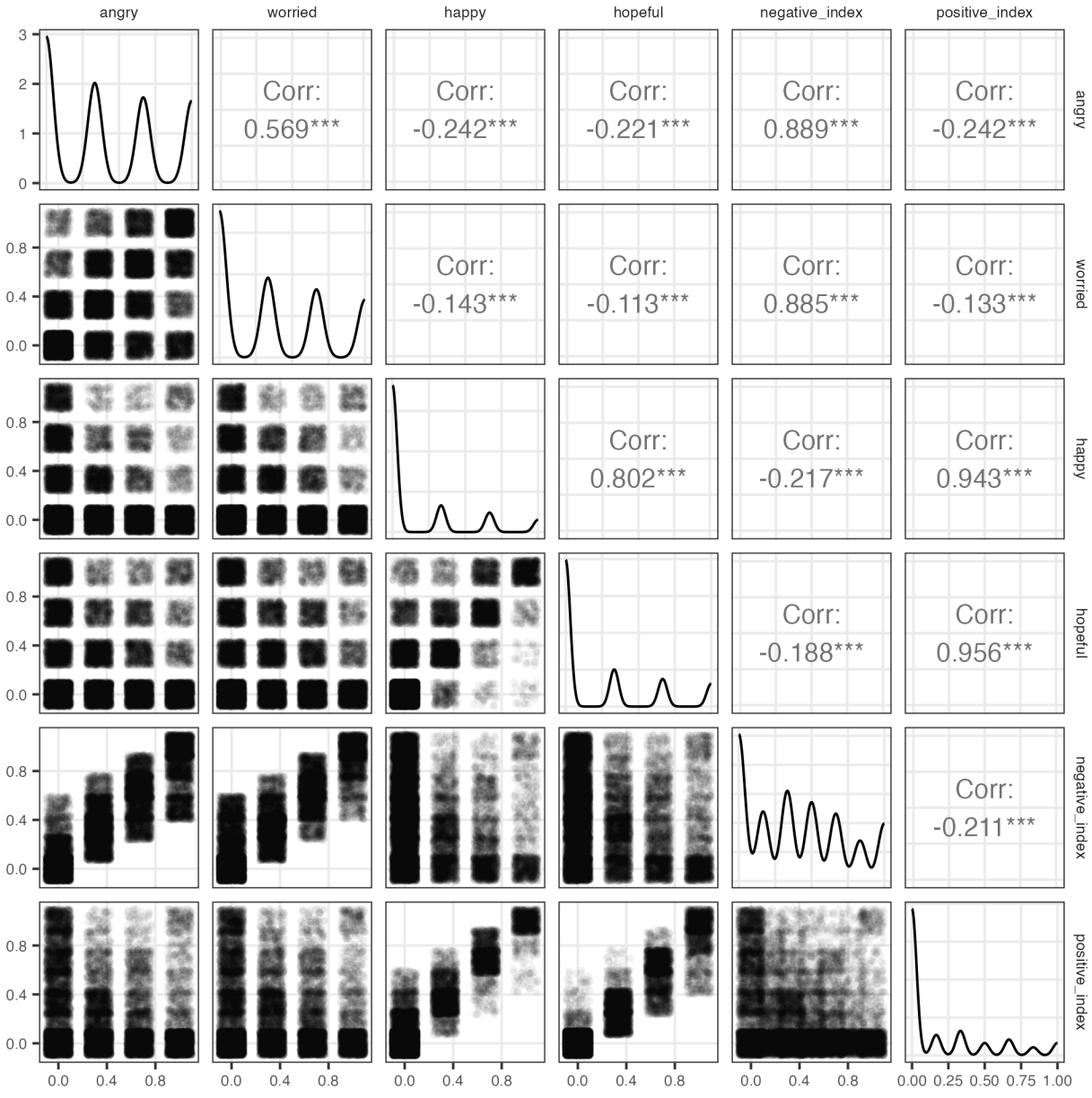


Figure A.2 shows the distributions and correlations among the emotion index components and the indicies.

Figure A.2: ggpairs Plot for Emotions and Emotional Indices



B Subgroup Ad Emotional Responses

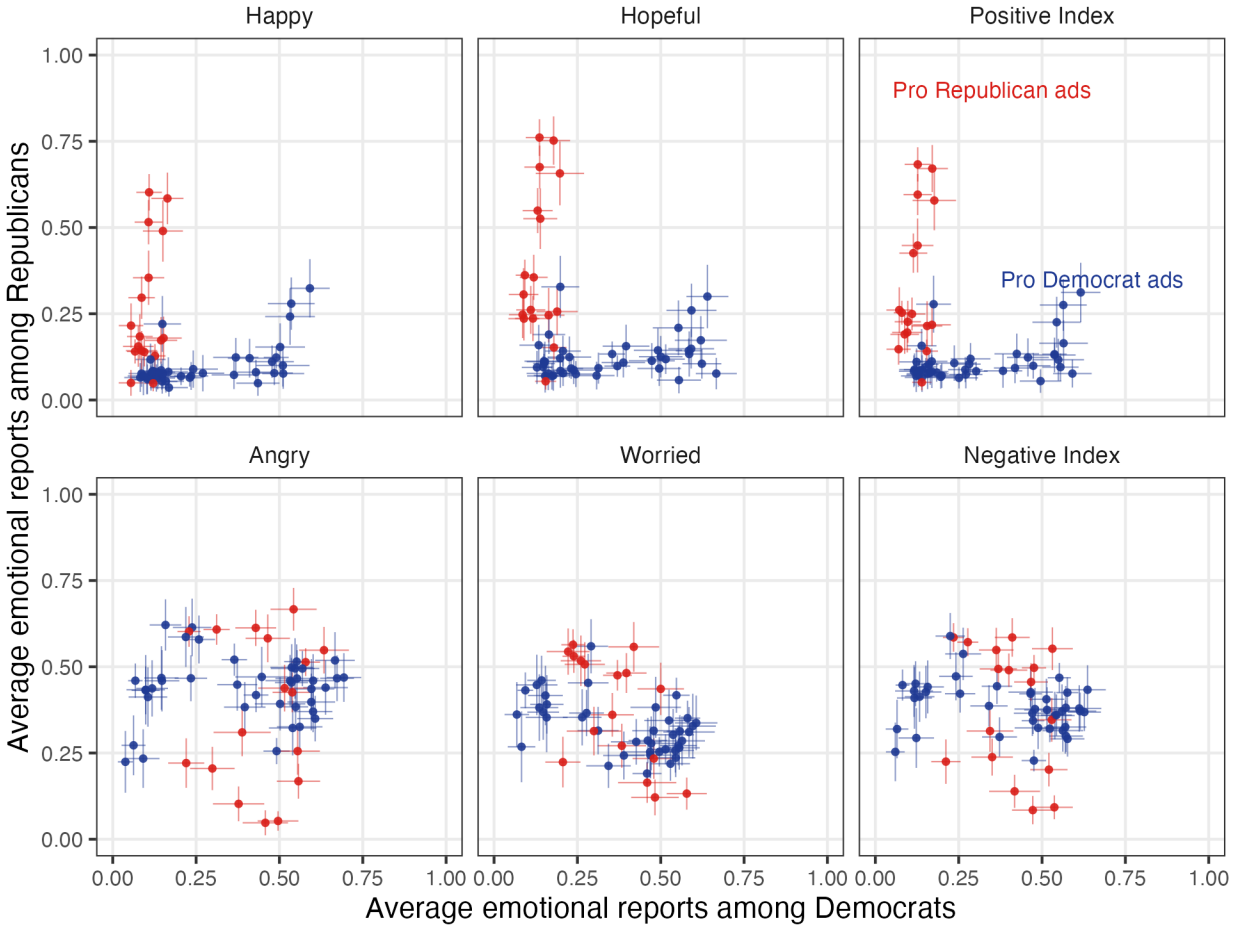


Figure B.3: Average emotion reports among Republicans and Democrats

C Metaregression tables

In this section, we present a series of regression tables. Each table is structured the same, with the first 6 columns subsetting the regression by respondent partisanship (Democrats, Independents, and Republicans) and by ad valence (Pro-Democrat (PD) or Pro-Republican (PR)). The seventh column pools all six subgroups together, and the eighth column includes controls for respondent partisanship and ad valence. We'll present the correlations with the positive emotion index on five outcome variables: candidate favorability, vote choice, turnout intentions, and target issue support (strict coding), and target issue support (stretch coding). We then turn to the same five tables with respect to the negative emotion index. The section concludes with a recreation of Figure 4 using target issue support (stretch coding) rather than target issue support (strict coding) as well as intended emotions.

C.1 Positive emotion index

Table C.1: Predicting Favorability CATEs from Positive Emotions

	Dem (PD)	Dem (PR)	Ind (PD)	Ind (PR)	Rep (PD)	Rep (PR)	All	All (2)
Intercept	0.034*	−0.012	0.013	−0.078	0.014	0.003	0.007	0.018*
	(0.011)	(0.050)	(0.026)	(0.050)	(0.011)	(0.012)	(0.005)	(0.008)
Positive Index	−0.039	0.026	0.000	0.381	−0.030	−0.006	0.008	−0.005
	(0.025)	(0.453)	(0.087)	(0.220)	(0.047)	(0.036)	(0.014)	(0.017)
Pro Republican								−0.017
								(0.009)
Independent								−0.012
								(0.013)
Republican								−0.004
								(0.007)
Observations	41	17	41	17	41	17	174	174
Q-test p	0.156	0.297	0.017	0.516	0.724	0.326	0.064	0.081

* p <0.05

Table C.2: Predicting Vote Choice CATEs from Positive Emotions

	Dem (PD)	Dem (PR)	Ind (PD)	Ind (PR)	Rep (PD)	Rep (PR)	All	All (2)
Intercept	0.005	0.039	0.122	0.015	0.064	−0.009	−0.001	0.015
	(0.019)	(0.052)	(0.068)	(0.062)	(0.037)	(0.013)	(0.006)	(0.010)
Positive Index	0.077	−0.532	−0.983	0.078	−0.523	−0.030	0.022	0.020
	(0.059)	(0.432)	(0.499)	(0.252)	(0.303)	(0.029)	(0.021)	(0.025)
Pro Republican								−0.033*
								(0.014)
Independent								0.001
								(0.014)
Republican								−0.005
								(0.010)
Observations	21	13	21	13	21	13	102	102
Q-test p	0.065	0.461	0.606	0.482	0.292	0.455	0.106	0.217

* p <0.05

Table C.3: Predicting Turnout Intentions CATEs from Positive Emotions

	Dem (PD)	Dem (PR)	Ind (PD)	Ind (PR)	Rep (PD)	Rep (PR)	All	All (2)
Intercept	−0.040	−0.013	0.105	0.006	−0.012	0.008	−0.017	−0.008
	(0.019)	(0.032)	(0.067)	(0.044)	(0.028)	(0.017)	(0.009)	(0.012)
Positive Index	0.161*	0.048	−0.974	−0.003	0.015	−0.017	0.064	0.060
	(0.057)	(0.246)	(0.512)	(0.279)	(0.297)	(0.036)	(0.035)	(0.033)
Pro Republican								−0.010
								(0.007)
Independent								−0.012
								(0.013)
Republican								−0.008
								(0.009)
Observations	21	13	21	13	21	13	102	102
Q-test p	0.548	0.701	0.892	0.563	0.557	0.697	0.790	0.769

* p <0.05

Table C.4: Predicting Target Issue CATEs from Positive Emotions

	Dem (PD)	Dem (PR)	Ind (PD)	Ind (PR)	Rep (PD)	Rep (PR)	All	All (2)
Intercept	0.011	0.000	−0.082	−0.090	0.003	0.057	0.010	−0.007
	(0.024)	(0.059)	(0.098)	(0.051)	(0.022)	(0.035)	(0.010)	(0.013)
Positive Index	−0.072	−0.110	0.609	0.386	0.179	−0.100	−0.039	−0.014
	(0.081)	(0.623)	(0.757)	(0.226)	(0.197)	(0.086)	(0.041)	(0.046)
Pro Republican								0.001
								(0.013)
Independent								0.001
								(0.019)
Republican								0.031*
								(0.010)
Observations	24	10	24	10	24	10	102	102
Q-test p	0.499	0.552	0.103	0.489	0.513	0.057	0.174	0.214

* p <0.05

Table C.5: Predicting Target Issue (Stretch Coding) CATEs from Positive Emotions

	Dem (PD)	Dem (PR)	Ind (PD)	Ind (PR)	Rep (PD)	Rep (PR)	All	All (2)
Intercept	0.012	−0.047	−0.058	−0.101*	−0.002	0.024	0.007	0.003
	(0.012)	(0.013)	(0.030)	(0.031)	(0.012)	(0.022)	(0.007)	(0.008)
Positive Index	−0.051	0.348	0.344	0.594*	0.172	−0.072	−0.025	−0.018
	(0.047)	(0.104)	(0.173)	(0.157)	(0.086)	(0.039)	(0.032)	(0.033)
Pro Republican								−0.005
								(0.008)
Independent								−0.006
								(0.010)
Republican								0.011
								(0.008)
Observations	55	23	55	23	55	23	234	234
Q-test p	0.156	0.625	0.189	0.271	0.619	0.438	0.076	0.081

* p <0.05

C.2 Negative emotion index

Table C.6: Predicting Favorability CATEs from Negative Emotions

	Dem (PD)	Dem (PR)	Ind (PD)	Ind (PR)	Rep (PD)	Rep (PR)	All	All (2)
Intercept	0.005	0.015	−0.021	−0.118	−0.020	0.021	0.004	0.007
	(0.006)	(0.032)	(0.036)	(0.113)	(0.027)	(0.027)	(0.005)	(0.006)
Negative Index	0.052*	−0.060	0.114	0.302	0.079	−0.046	0.014	0.034
	(0.023)	(0.102)	(0.126)	(0.288)	(0.075)	(0.050)	(0.017)	(0.020)
Pro Republican								−0.020*
								(0.009)
Independent								−0.010
								(0.013)
Republican								−0.006
								(0.006)
Observations	41	17	41	17	41	17	174	174
Q-test p	0.217	0.314	0.018	0.383	0.755	0.359	0.066	0.100

* p <0.05

Table C.7: Predicting Vote Choice CATEs from Negative Emotions

	Dem (PD)	Dem (PR)	Ind (PD)	Ind (PR)	Rep (PD)	Rep (PR)	All	All (2)
Intercept	0.048	−0.053	−0.139	−0.020	−0.028	−0.035	0.006	0.014
	(0.031)	(0.050)	(0.185)	(0.099)	(0.096)	(0.016)	(0.017)	(0.015)
Negative Index	−0.049	0.065	0.377	0.118	0.098	0.039	−0.005	0.015
	(0.064)	(0.122)	(0.500)	(0.267)	(0.228)	(0.032)	(0.041)	(0.036)
Pro Republican								−0.033*
								(0.013)
Independent								−0.001
								(0.014)
Republican								−0.007
								(0.010)
Observations	21	13	21	13	21	13	102	102
Q-test p	0.049	0.391	0.373	0.478	0.239	0.456	0.101	0.215

* p <0.05

Table C.8: Predicting Turnout Intentions CATEs from Negative Emotions

	Dem (PD)	Dem (PR)	Ind (PD)	Ind (PR)	Rep (PD)	Rep (PR)	All	All (2)
Intercept	0.049	−0.047	0.090	−0.028	0.013	−0.012	0.004	0.012
	(0.030)	(0.079)	(0.111)	(0.159)	(0.044)	(0.017)	(0.010)	(0.013)
Negative Index	−0.097	0.092	−0.286	0.080	−0.061	0.034	−0.017	−0.014
	(0.062)	(0.162)	(0.295)	(0.385)	(0.102)	(0.037)	(0.021)	(0.024)
Pro Republican								−0.009
								(0.007)
Independent								−0.018
								(0.014)
Republican								−0.010
								(0.009)
Observations	21	13	21	13	21	13	102	102
Q-test p	0.290	0.754	0.752	0.566	0.573	0.710	0.692	0.681

* p <0.05

Table C.9: Predicting Target Issue CATEs from Negative Emotions

	Dem (PD)	Dem (PR)	Ind (PD)	Ind (PR)	Rep (PD)	Rep (PR)	All	All (2)
Intercept	−0.051	−0.321	0.011	0.032	0.117	−0.018	−0.020	−0.046
	(0.027)	(0.110)	(0.101)	(0.131)	(0.097)	(0.039)	(0.028)	(0.025)
Negative Index	0.091	0.708	−0.032	−0.090	−0.238	0.096	0.052	0.078
	(0.055)	(0.258)	(0.276)	(0.315)	(0.229)	(0.083)	(0.059)	(0.045)
Pro Republican								0.001
								(0.012)
Independent								0.008
								(0.020)
Republican								0.035*
								(0.009)
Observations	24	10	24	10	24	10	102	102
Q-test p	0.558	0.912	0.071	0.394	0.535	0.056	0.175	0.262

* p <0.05

Table C.10: Predicting Target Issue (Stretch Coding) CATEs from Negative Emotions

	Dem (PD)	Dem (PR)	Ind (PD)	Ind (PR)	Rep (PD)	Rep (PR)	All	All (2)
Intercept	−0.017	−0.076	−0.027	0.229	0.117*	−0.039	−0.012	−0.015
	(0.023)	(0.048)	(0.042)	(0.103)	(0.044)	(0.018)	(0.014)	(0.017)
Negative Index	0.035	0.178	0.043	−0.513	−0.256*	0.092	0.032	0.032
	(0.046)	(0.105)	(0.134)	(0.226)	(0.105)	(0.049)	(0.033)	(0.034)
Pro Republican								−0.005
								(0.007)
Independent								−0.002
								(0.012)
Republican								0.014
								(0.008)
Observations	55	23	55	23	55	23	234	234
Q-test p	0.136	0.695	0.117	0.136	0.738	0.453	0.079	0.089

* p <0.05

C.3 Intended emotion index

Table C.11: Predicting Favorability CATEs from Intended Emotions

	Dem (PD)	Dem (PR)	Ind (PD)	Ind (PR)	Rep (PD)	Rep (PR)	All	All (2)
Intercept	0.020	−0.003	0.000	−0.088*	0.012*	0.034	0.010*	0.018*
	(0.022)	(0.019)	(0.022)	(0.032)	(0.004)	(0.032)	(0.004)	(0.007)
Intended Emotion Index	0.002	−0.029	0.066	0.228*	−0.035	−0.073	−0.004	−0.005
	(0.049)	(0.055)	(0.077)	(0.057)	(0.017)	(0.056)	(0.013)	(0.013)
Pro Republican								−0.016
								(0.008)
Independent								−0.012
								(0.013)
Republican								−0.004
								(0.007)
Observations	41	17	41	17	41	17	174	174
Q-test p	0.118	0.330	0.031	0.613	0.827	0.456	0.064	0.082

* p <0.05

Table C.12: Predicting Vote Choice CATEs from Intended Emotions

	Dem (PD)	Dem (PR)	Ind (PD)	Ind (PR)	Rep (PD)	Rep (PR)	All	All (2)
Intercept	0.034	−0.026	−0.031	−0.020	0.010	0.005	0.009	0.021*
	(0.023)	(0.021)	(0.028)	(0.087)	(0.013)	(0.032)	(0.005)	(0.007)
Intended Emotion Index	−0.029	0.004	0.178	0.127	0.008	−0.047	−0.023	−0.002
	(0.066)	(0.058)	(0.115)	(0.190)	(0.037)	(0.053)	(0.024)	(0.021)
Pro Republican								−0.032*
								(0.013)
Independent								−0.001
								(0.014)
Republican								−0.006
								(0.011)
Observations	21	13	21	13	21	13	102	102
Q-test p	0.046	0.376	0.434	0.506	0.228	0.497	0.120	0.212

* p <0.05

Table C.13: Predicting Turnout Intentions CATEs from Intended Emotions

	Dem (PD)	Dem (PR)	Ind (PD)	Ind (PR)	Rep (PD)	Rep (PR)	All	All (2)
Intercept	0.032	−0.006	−0.016	0.035	−0.013	0.002	−0.001	0.007
	(0.013)	(0.009)	(0.019)	(0.082)	(0.007)	(0.033)	(0.006)	(0.007)
Intended Emotion Index	−0.083	−0.004	0.016	−0.075	0.030	0.001	−0.007	−0.003
	(0.038)	(0.015)	(0.123)	(0.262)	(0.023)	(0.057)	(0.012)	(0.011)
Pro Republican								−0.009
								(0.007)
Independent								−0.017
								(0.014)
Republican								−0.010
								(0.009)
Observations	21	13	21	13	21	13	102	102
Q-test p	0.357	0.700	0.709	0.577	0.608	0.690	0.691	0.677

* p <0.05

Table C.14: Predicting Target Issue CATEs from Intended Emotions

	Dem (PD)	Dem (PR)	Ind (PD)	Ind (PR)	Rep (PD)	Rep (PR)	All	All (2)
Intercept	0.000	−0.040	0.029	−0.013	0.017	0.035	0.000	−0.014
	(0.015)	(0.021)	(0.034)	(0.084)	(0.012)	(0.037)	(0.009)	(0.009)
Intended Emotion Index	−0.032	0.091	−0.138	0.025	0.027	−0.024	0.009	0.017
	(0.056)	(0.060)	(0.140)	(0.211)	(0.037)	(0.062)	(0.023)	(0.018)
Pro Republican								−0.002
								(0.012)
Independent								0.003
								(0.019)
Republican								0.033*
								(0.009)
Observations	24	10	24	10	24	10	102	102
Q-test p	0.434	0.796	0.090	0.394	0.502	0.049	0.152	0.216

* p <0.05

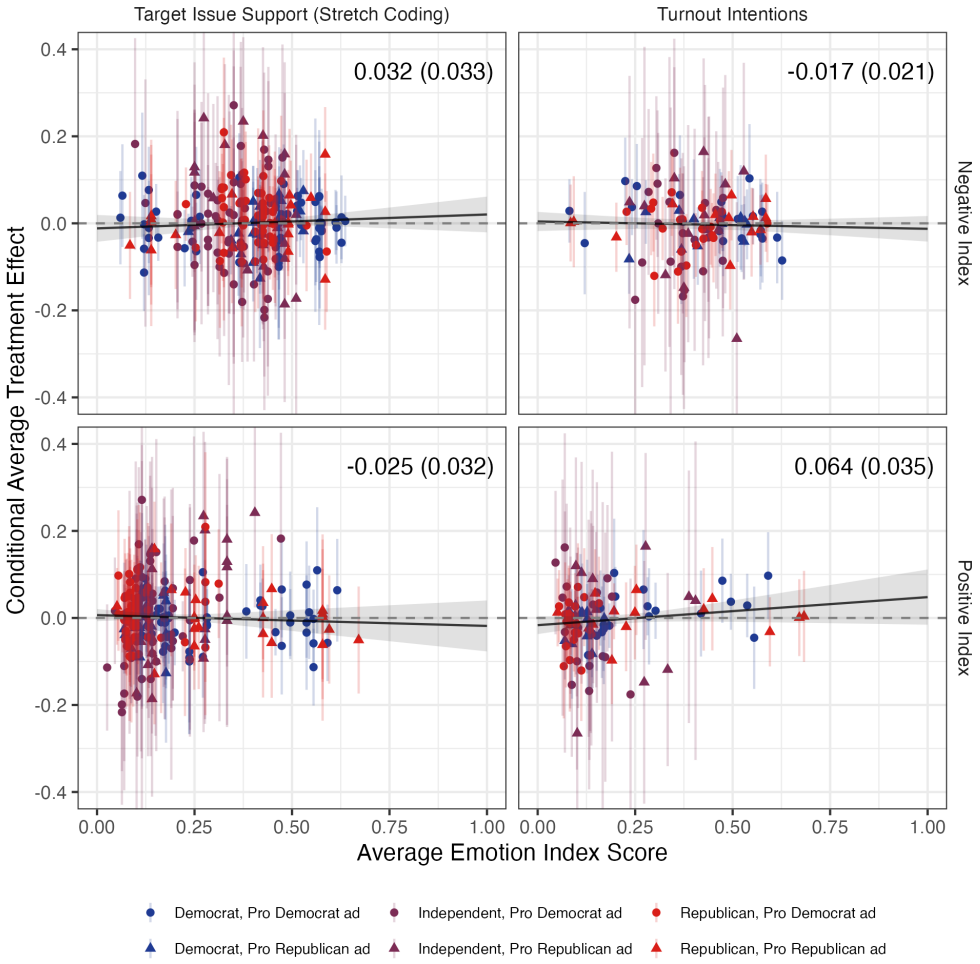
Table C.15: Predicting Target Issue (Stretch Coding) CATEs from Intended Emotions

	Dem (PD)	Dem (PR)	Ind (PD)	Ind (PR)	Rep (PD)	Rep (PR)	All	All (2)
Intercept	0.004	−0.006	−0.020	0.007	0.011	−0.002	0.000	−0.005
	(0.012)	(0.007)	(0.009)	(0.057)	(0.007)	(0.027)	(0.003)	(0.005)
Intended Emotion Index	−0.017	0.006	0.052	0.041	0.028	0.006	0.007	0.013
	(0.036)	(0.024)	(0.064)	(0.112)	(0.022)	(0.037)	(0.010)	(0.010)
Pro Republican								−0.006
								(0.007)
Independent								−0.003
								(0.011)
Republican								0.015
								(0.007)
Observations	55	23	55	23	55	23	234	234
Q-test p	0.121	0.517	0.127	0.055	0.621	0.386	0.069	0.082

* p <0.05

C.4 Remaking Figure 4 with Target Issue (Stretch Coding)

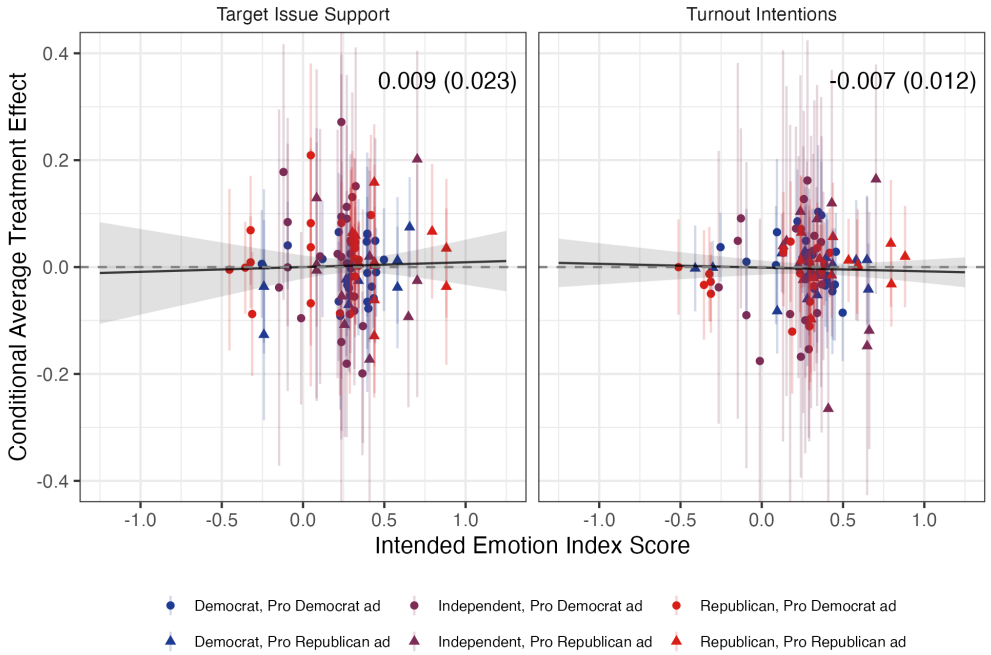
Figure C.4: Predicting Turnout Intentions and Target Issue Support (Stretch Coding) CATEs from Emotions



Note. Red points correspond to an ad's average emotion index score and CATE among Republican respondents; blue points among Democrats; and purple points among independents. The points are plotted with triangles for anti-Clinton or pro-Trump ads and with circles for pro-Clinton or anti-Trump ads.

C.5 Remaking Figure 4 with Intended Emotions

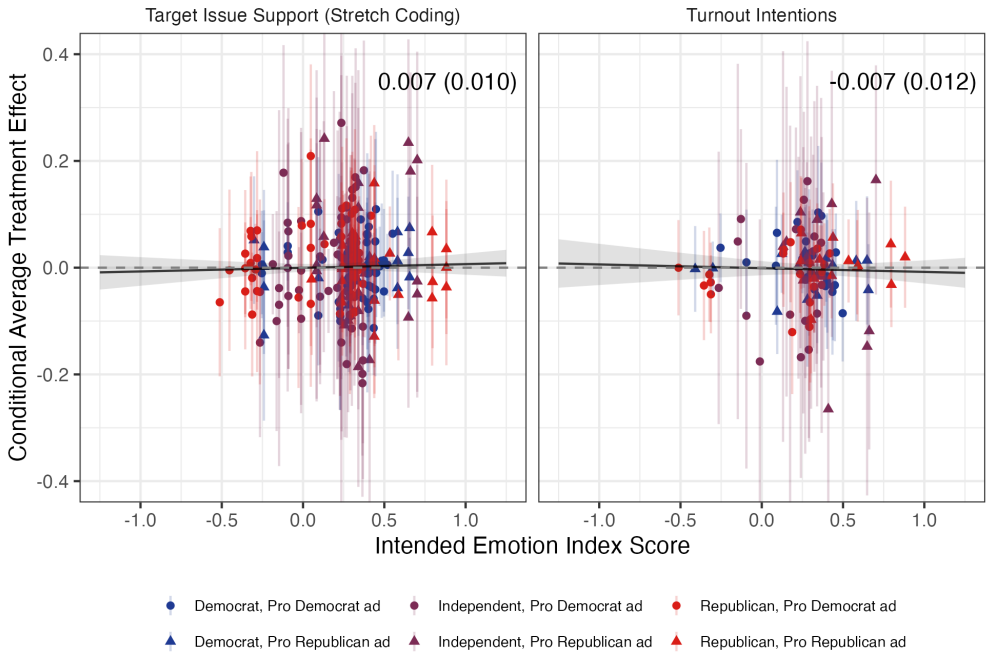
Figure C.5: Predicting Turnout Intentions and Target Issue Support CATEs from Intended Emotions



Note. Red points correspond to an ad's average emotion index score and CATE among Republican respondents; blue points among Democrats; and purple points among independents. The points are plotted with triangles for anti-Clinton or pro-Trump ads and with circles for pro-Clinton or anti-Trump ads.

C.6 Remaking Figure 4 with Target Issue (Stretch Coding) and Intended Emotions

Figure C.6: Predicting Turnout Intentions and Target Issue Support (Stretch Coding) CATEs from Intended Emotions



Note. Red points correspond to an ad's average emotion index score and CATE among Republican respondents; blue points among Democrats; and purple points among independents. The points are plotted with triangles for anti-Clinton or pro-Trump ads and with circles for pro-Clinton or anti-Trump ads.

D Intended Emotion Codings

Table D.16 shows our coding of the valence of each ad and also, in our estimation, the intended emotional valence of each add. An ad with a "positive" emotional valence is supposed to evoke more positive emotions and an ad with a "negative" emotional valence is supposed to evoke more negative emotions. These codings are instrumental for our "intended emotions" analysis in the main text.

Table D.16: Intended Emotion Coding for each Advertisement

Ad ID	Title	Valence	Emotional Valence
4	Scam	Anti Trump	negative
6	Plain Wrong	Anti Trump	negative
8	Quotes	Anti Trump	negative
9	Real Life	Pro Clinton	positive
11	America	Pro Sanders	positive
12	All the Good	Pro Clinton	positive
17	Erica	Pro Sanders	both
18	I'm with him	Pro Clinton	positive
20	Math	Pro Cruz	positive
21	Suggest	Pro Kasich	both
23	Not Easy	Pro Cruz	negative
24	Broken	Pro Trump	positive
26	Unifier	Anti Trump	negative
27	Harrassment	Anti Trump	negative
30	Love/Kindness	Pro Clinton	positive
33	Speak	Anti Trump	negative
38	Extreme Makeover	Anti Trump	negative
39	Grace	Anti Trump	negative
42	Moments	Pro Clinton	positive
44	Tested	Anti Trump	negative
45	Stop Clinton	Anti Clinton	negative
47	Role Model	Anti Trump	negative
48	Same Old	Anti Clinton	negative
50	Someplace	Anti Trump	negative
51	Outsource	Anti Clinton	negative
54	Unfit	Anti Trump	negative
56	Depend	Pro Clinton	positive
57	2 Americas: Immigration	Anti Clinton	both
58	2 Americas: Economy	Anti Clinton	both
59	Everything	Anti Trump	negative
60	Sacrifice	Anti Trump	negative
61	Deplorables	Anti Clinton	negative
62	Agree	Anti Trump	negative
63	Mirrors	Anti Trump	negative
64	Defenseless	Anti Clinton	negative
65	50-pts?	Anti Clinton	negative
66	Arrogant	Anti Trump	negative
67	Change	Anti Clinton	both
68	America's Bullies	Anti Trump	positive
69	Laura	Anti Clinton	negative
70	Captain Khan	Anti Trump	negative
71	Example	Anti Trump	positive
72	Consumer Benefit	Pro Trump	positive
74	Movement	Pro Trump	positive
75	Families	Anti Trump	positive
76	Unfit to Serve	Anti Clinton	negative
77	Daughters	Anti Trump	negative
99	Sacrifice	Anti Trump	negative

E Additional Ratings Analysis

In this section, we discuss the correlation of subjects' subjective ratings of the persuasiveness of advertisements with the actual persuasiveness as measured by the survey questions. *Rating questions* are survey measures that political scientists and political professionals sometimes use when trying to evaluate the relative persuasiveness of political advertisements without running the survey experiments. The hope of course is that the ratings will serve as a guide to the persuasiveness of the ad, both on average and across individuals. Ads that receive higher ratings should generate larger average treatment effects on vote choice and candidate evaluations. At an individual level, those who rate ads more highly should experience larger treatment effects than those who give lower ratings.

The use of rating questions for the purpose of predicting persuasiveness is widespread. For example, political strategists commonly use focus groups to gauge the persuasiveness of advertisements. In their comprehensive detailing of the rise of the role of political consultants, Thurber and Nelson (2001) write that focus groups are a central part of the modern campaign and that it is commonplace for political consultants to turn to focus groups in order to conduct candidate research. In a typical focus group, "[a] moderator might show an ad, pass around a brochure, or read a snippet from a possible speech, and then call on people around the table to see what they thought of it" (Issenberg 2012, p. 152). Issenberg further documents that hosting nightly focus groups in key battleground states became a frequent practice of the 2008 Obama campaign.

Political practitioners, however, are not alone in their use of subjective ratings to approximate persuasion. Some political scientists use subjective persuasion as an accompanying measure to persuasion, while others directly infer effectiveness from subjective ratings. A 2012 study on the role of source credibility in persuasion employs measures of both "subjective persuasion," in which subjects reported how persuaded and convinced they were by the ad, and "objective persuasion," in which subjects were asked their favorability and vote choice for a hypothetical candidate Weber, Dunaway, and Johnson (2012). In contrast, in order to measure an ad's effectiveness, Searles et al. (2020)) measure an ad's credibility, informativeness, and ability to grab one's attention. They do this by directly asking respondents how credible the ad was, how much they learned from the ad, and how much it grabbed their attention.

We are not the first to examine the relationship between ratings and persuasiveness. In their first study, Blumenau and Lauderdale (2024) ask subjects to rate 336 treatments. In the second study, they measure the persuasiveness of the three best-rated and three worst-rated ads (or a control) on 12 issues via randomized controlled trial, finding indeed that the best-rated arguments yielded larger treatment effects than the worst-rated arguments. An important feature of this study is the 336 treatments were created by the researchers by fully crossing a set of possible rhetorical styles, so it is possible that the worst-rated treatments were especially bizarre or unlikely to occur in a real campaign. This criticism does not challenge the finding that ratings in that study predicted persuasiveness; instead it raises for us the question of whether ratings of real campaign advertisements would perform as well.

O'Keefe (2018) takes up the same question of whether subjective ratings are, on average, related to persuasiveness in a review of psychology studies. O'Keefe identifies a large number of psychology studies that have measures of both perceived effectiveness and actual effectiveness of pairs of treatments. He finds that in 151 such comparisons, ads perceived to be more effective are actually more effective 58% of the time. As this estimate was not significantly different from 50%,

O'Keefe (2018) concludes, "[a]cross all cases, pretesting messages by asking respondents about perceived or expected persuasiveness was no more informative about relative actual persuasiveness than flipping a coin"(O'Keefe 2018, p. 133).

In our own studies, subjects were first asked to rate the ad's effectiveness: "On a scale of 1 to 100, where would you place this ad in terms of overall quality?" To provide structure to the responses, subjects were told that "1 = Well below average, among the worst campaign ads I've ever seen", and "100 = Well above average, one of the best campaign ads I've ever seen."

Next, subjects were told, "We'd like you to think like an advertising expert and consider the effectiveness of this ad. We are interested in your thoughts even if you don't think you would ever vote for this candidate." They were then asked, "Based on your assessment of this ad, if you were advising the campaign, how often would you suggest they run this ad?" But before they provided a responses, they were reminded, "Remember, we are asking you to take your politics out of it and answer as if you were hired as an advisor to this candidate." Responses were measured on a three-point scale from "1 = I'd run this ad a lot, it's good at accomplishing its goal", to "3 = I wouldn't run it at all, it just doesn't work." For the final three rating questions, subjects were asked to rate the ad's fairness, truthfulness, and memorability on a five-point scale from "very [rating]" to "very un[rating]".

We combine these five ratings into an index of the five ratings questions. We standardized the index components to fall in the 0 to 1 range, then took the average. The resulting index is highly reliable (Cronbach's alpha: 0.91). Figure E.7 shows the distributions of and correlations among the five ratings questions and the additive index.

As shown in Figure E.8, all the ratings load on a single factor. We take these results to mean that despite the seeming differences in the ratings questions (e.g., fairness and memorability are quite different concepts), subjects used all the ratings questions to express a common sentiment towards the ads, which we interpret as their affective evaluation of the ads.

In Figure E.9, we show that the average ratings by subject partisanship show a very clear pattern: In-partisans rate partisan-congenial ads more highly than out-partisans, regardless of the measure we use.

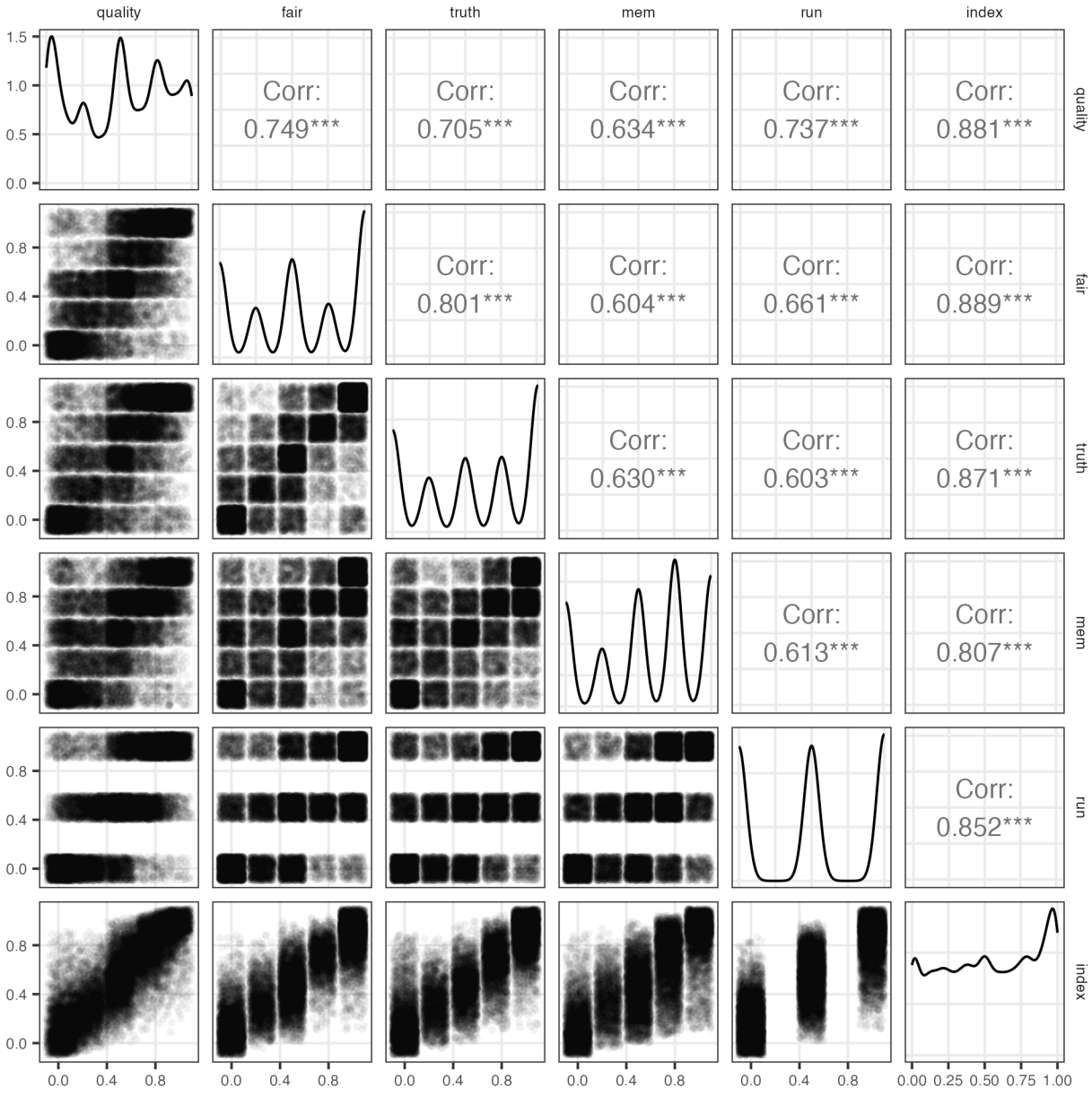
Figure E.10 displays the relationship subjects' subjective ratings with effectiveness. The estimated meta-regression slopes for the favorability and vote choice outcomes are both positive, with the slope for favorability estimated to be statistically significant. The estimated slope of 0.034 indicates that an ad achieving the maximum rating of 1 relative to the minimum rating of 0 would be 3.4 points higher.

Lastly, we present regression tables that describe the association between the rating index and the effects on favorability and vote choice CATEs, separately by partisanship (Democrats, Independents, and Republicans) and the valence of the ads (Pro-Democratic or Pro-Republican). Three of 12 slopes are positive and statistically distinct from zero. We also report an analysis that controls for respondent partisanship and ad valence, finding a similar result to the unconditional estimates.

E.1 Assessing pairwise prediction accuracy

In this section, we conduct an analysis akin to O'Keefe (2018) that tests how often more highly rated advertisements are more persuasive than those rated more lowly, at the group and individual levels. At the individual-level we analyze how accurate individual ratings are at predicting the relative persuasiveness of ads for both favorability and vote choice. Our design allows us to analyze

Figure E.7: Correlations among the Five Rating Index Variables



individual accuracy by subsetting our sample to individuals who viewed two advertisements (n = 5,070); among these subjects we can estimate how frequently they rate the more persuasive advertisement more highly. Estimates of accuracy are benchmarked against 0.50 (random guessing).

We present overall estimates and estimates by matchup: a comparison of two pro-Democratic ads, two pro-Republican ads, or a pro-Democratic ad versus a pro-Republican ad. We also conduct the same analysis by respondent partisanship. Because our estimates of ratings and CATEs are not independent, we generate uncertainty estimates via the nonparametric bootstrap with 1,000 resamples.

Figure E.11 plots the results of the individual-level pairwise analysis. Each point is the pro-

Figure E.8: Scree Plot for Rating Index Factors

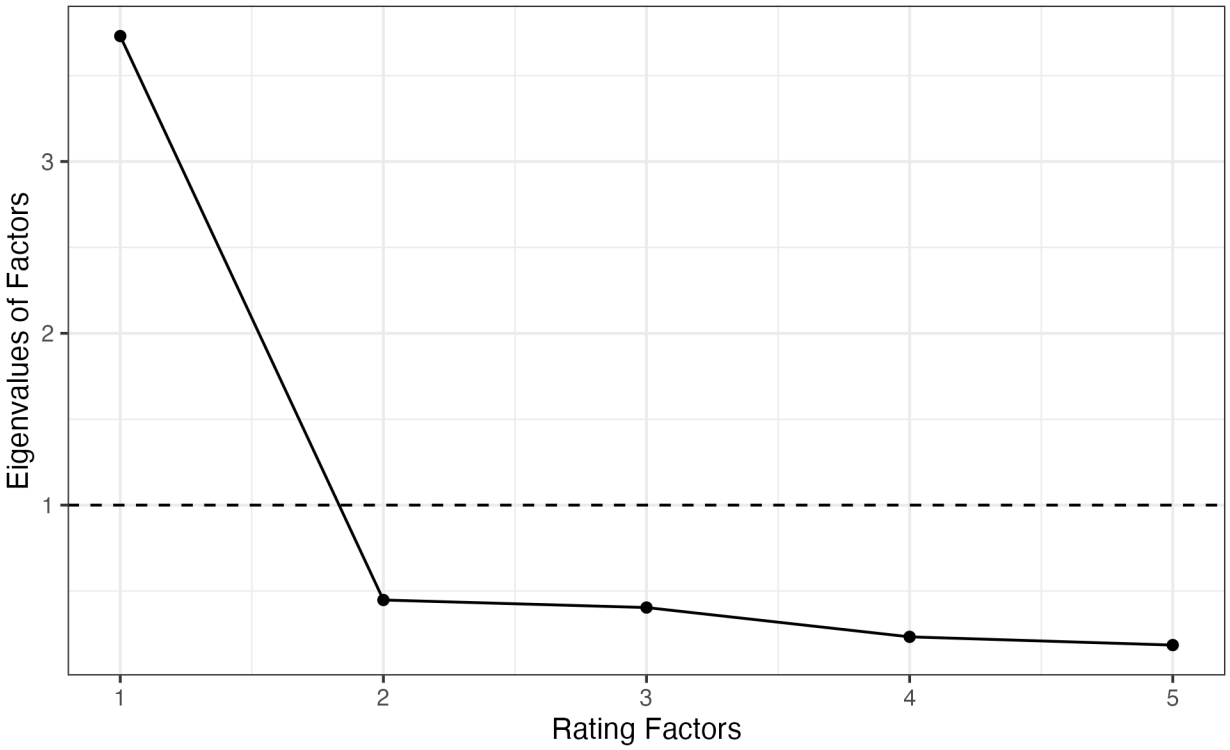


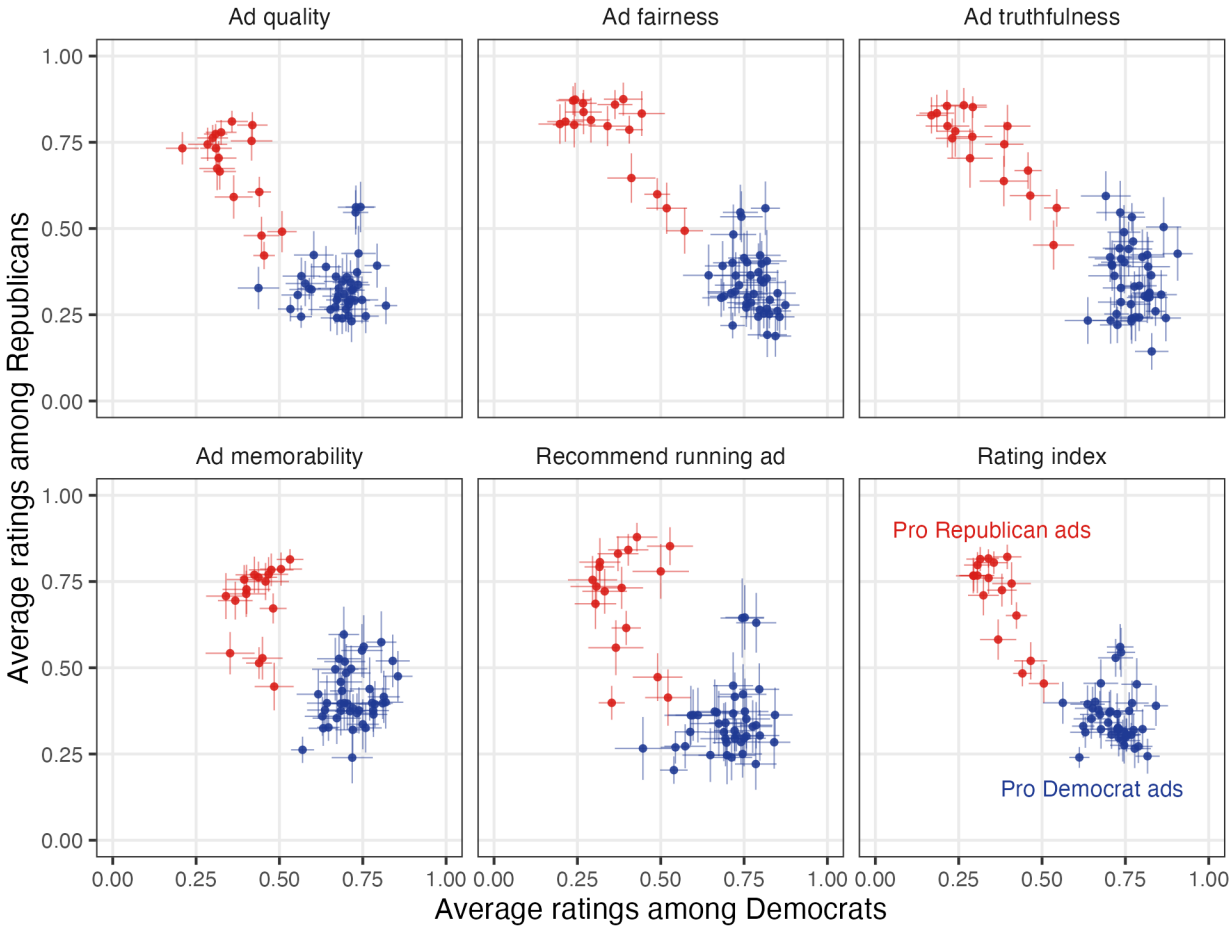
Table E.17: Predicting Favorability CATEs from Ratings

	Dem (PD)	Dem (PR)	Ind (PD)	Ind (PR)	Rep (PD)	Rep (PR)	All	All (2)
Intercept	−0.188*	−0.087	−0.093	−0.414*	0.030	0.054	−0.010	−0.007
	(0.082)	(0.073)	(0.050)	(0.115)	(0.019)	(0.056)	(0.009)	(0.011)
Rating Index	0.292*	0.209	0.223	0.780*	−0.051	−0.075	0.034*	0.036
	(0.117)	(0.177)	(0.104)	(0.227)	(0.039)	(0.078)	(0.016)	(0.019)
Pro Republican								−0.018
								(0.009)
Independent								−0.005
								(0.012)
Republican								0.002
								(0.006)
Observations	41	17	41	17	41	17	174	174
Q-test p	0.225	0.367	0.035	0.758	0.745	0.370	0.083	0.100

* p <0.05

portion of individuals within each condition and partisan group who viewed two advertisements and rated the more effective (persuasive) ad more highly when asked to do so subjectively. "More effective" is measured within partisan subgroup. The results in Figure E.11 suggest that, overall,

Figure E.9: Average ad ratings among Republicans and Democrats

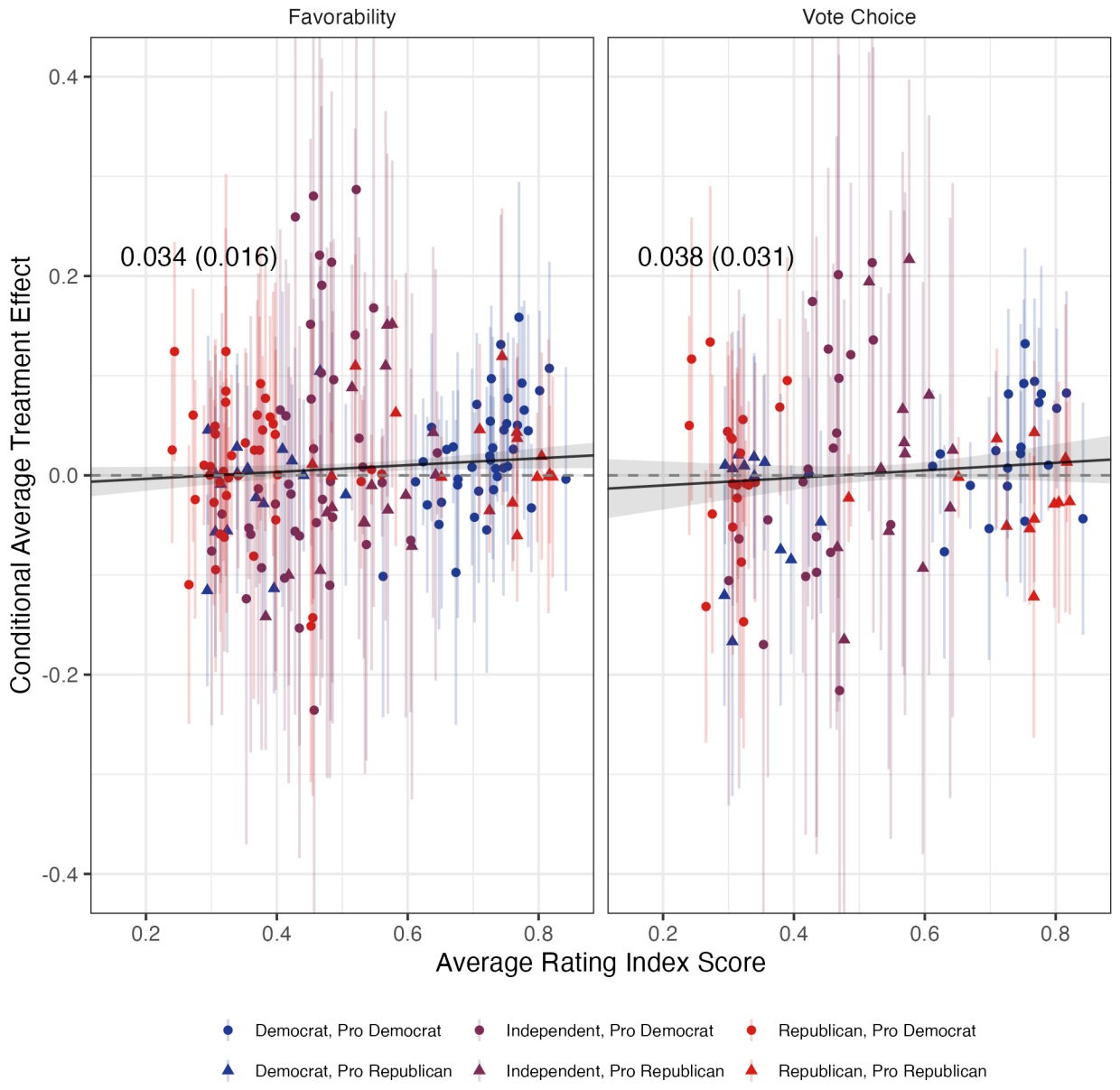


individuals are no better than random guessing at accurately predicting which ad will be more persuasive.

At first glance, Democrats appear to be much better than Republicans at identifying the more persuasive advertisement. However, this pattern is spurious because partisans tend to rate in-party ads more highly and the set of Democratic ads in the sample are slightly more effective. For example, in pro-Democrat vs. pro-Republican advertisement matchups, the pro-Democrat ad had a larger CATE among Democrats 66.7% of the time for favorability and 83.3% of the time for vote choice. On the other hand, in the same matchups, for Republicans, pro-Republican ads had a larger CATE only 33.3% of the time for favorability and 33.3% of the time for vote choice. Interestingly, in the same matchups, the pro-Democratic and pro-Republican ads had larger CATES for favorability and vote choice in exactly the same percentage of cases: 50%. Partisans also show strong favoritism toward in-party advertisements: in pro-Democrat v. pro-Republican ad matchups, Democrats gave higher ratings to the pro-Democratic ad 84.5% of the time and Republicans gave higher ratings to the pro-Republican ad 87.1% of the time.

The group-level results in E.12 are substantively identical. The confidence intervals are much larger because the unit of study is not the individual, but the party subgroup viewing a pair of ads. For favorability, there are six cases in "Dem v. Rep," seven in "Dem v. Dem" and two in

Figure E.10: Predicting CATEs from Ratings



Note. Red points correspond to an ad's average emotion index score and CATE among Republican respondents; blue points among Democrats; and purple points among independents. The points are plotted with triangles for anti-Clinton or pro-Trump ads and with circles for pro-Clinton or anti-Trump ads.

"Rep v. Rep." For vote choice, there are six cases in "Dem v. Rep," one in "Dem v. Dem," and zero in "Rep v. Rep." In-party favoritism in ratings is even more pronounced in the group-level analysis: Average Democratic (Republican) ratings of pro-Democratic (pro-Republican) ads are higher than ratings of pro-Republican (pro-Democratic) ads 100% of the time. Independents,

Table E.18: Predicting Vote Choice CATEs from Ratings

	Dem (PD)	Dem (PR)	Ind (PD)	Ind (PR)	Rep (PD)	Rep (PR)	All	All (2)
Intercept	−0.192	0.002	−0.374*	−0.298	−0.026	−0.025	−0.018	−0.006
	(0.118)	(0.112)	(0.123)	(0.314)	(0.153)	(0.034)	(0.017)	(0.022)
Rating Index	0.300	−0.077	0.864*	0.592	0.112	0.013	0.038	0.044
	(0.164)	(0.298)	(0.310)	(0.564)	(0.480)	(0.056)	(0.031)	(0.037)
Pro Republican								−0.035*
								(0.013)
Independent								0.005
								(0.016)
Republican								−0.002
								(0.010)
Observations	21	13	21	13	21	13	102	102
Q-test p	0.126	0.381	0.588	0.558	0.231	0.435	0.122	0.260

* p <0.05

followed the pattern of Republicans and rated pro-Republican ads higher in all cases. As a result of this, the proportion of ads correctly classified in E.12 in the "Dem v. Rep" rows is exactly the fraction of times the in-party rated ad has a larger CATE. Democrats correctly predict "Dem v. Rep" favorability CATEs 66.7% of the time and vote choice CATEs 83.3% of the time because that is how often pro-Democratic ads had larger CATEs. Republicans correctly predict "Dem v. " favorability CATEs 33.3% of the time and vote choice CATEs 33.3% of the time because that is how often pro-Republican ads had larger CATEs. Independents correctly predict "Dem v. Rep" favorability CATEs 50% of the time and vote choice CATEs 50% of the time because that is how often pro-Republicans ads had larger CATEs. Vote choice cannot be evaluated for pro-Republican versus pro-Republican matchups, because both such matchups were fielded during the primary period, when the general election vote choice question was not asked.

Figure E.11: Individual-Level Accuracy Analysis

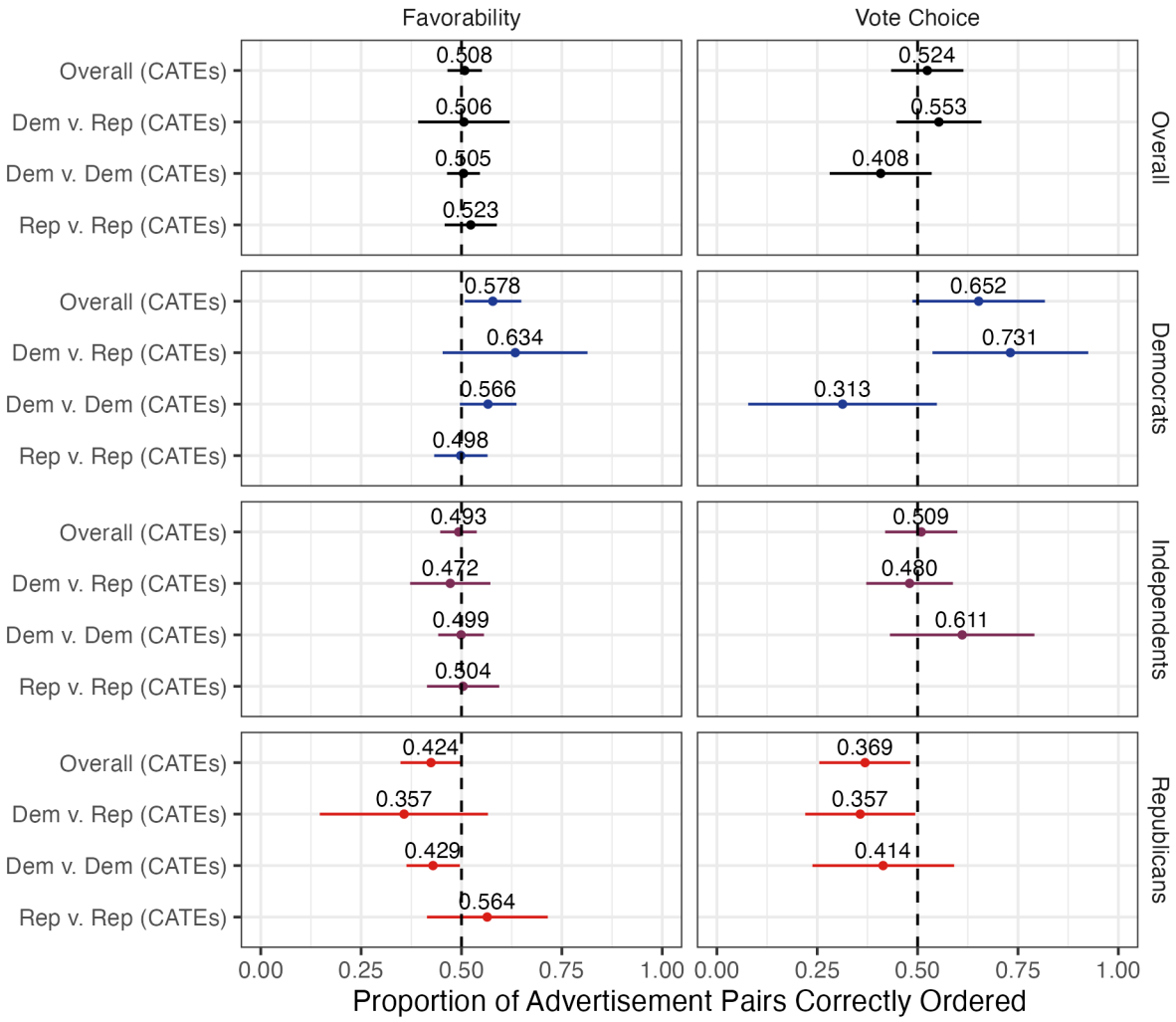
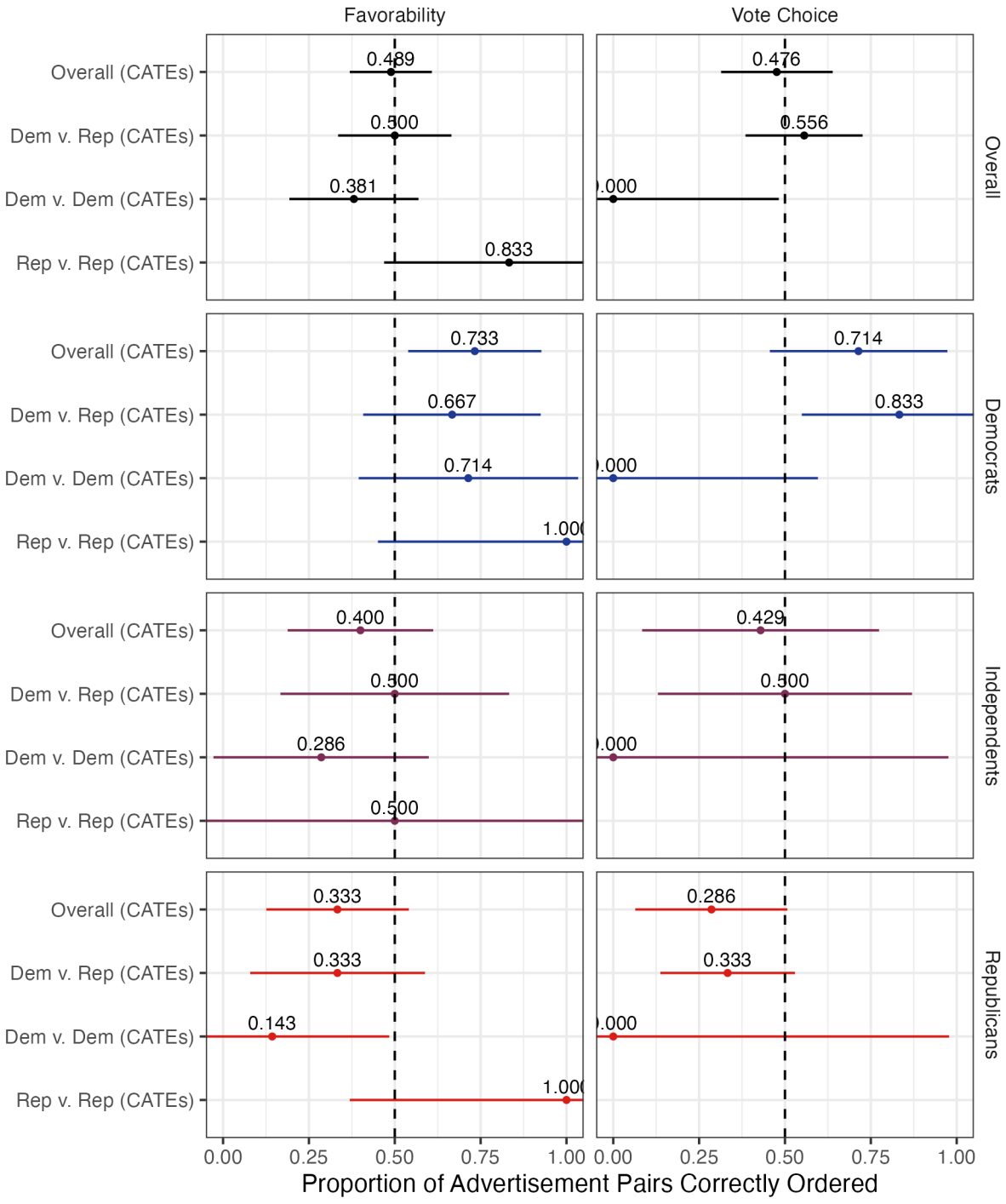


Figure E.12: Group-Level Accuracy Analysis



References

Blumenau, Jack, and Benjamin E Lauderdale. 2024. "The variable persuasiveness of political rhetoric." *American Journal of Political Science* 68 (1): 255–270.

Issenberg, Sasha. 2012. *The victory lab: The secret science of winning campaigns*. Crown.

O'Keefe, Daniel J. 2018. "Message pretesting using assessments of expected or perceived persuasiveness: Evidence about diagnosticity of relative actual persuasiveness." *Journal of Communication* 68 (1): 120–142.

Searles, Kathleen, Erika Franklin Fowler, Travis N Ridout, Patricia Strach, and Katherine Zuber. 2020. "The effects of men's and women's voices in political advertising." *Journal of Political Marketing* 19 (3): 301–329.

Thurber, James A., and Candice J. Nelson. 2001. *Campaign warriors: Political consultants in elections*. Rowman & Littlefield.

Weber, Christopher, Johanna Dunaway, and Tyler Johnson. 2012. "It's all in the name: Source cue ambiguity and the persuasive appeal of campaign ads." *Political Behavior* 34: 561–584.