Appendix for "Praise from Peers Promotes Empathetic Behavior"

Adeline Lo[*] Jonathan Renshon[†] Lotem Bassan-Nygate[‡]

[*]Assistant Professor of Political Science, UW-Madison.: aylo@wisc.edu
[†]Associate Professor of Political Science, UW-Madison. : renshon@wisc.com
[‡]Graduate student of Political Science, UW-Madison. : lbassan@wisc.edu

Contents

A Summarizing information on studies

	Study goal	Respondents	Trials	Total obs.	Treatment arms
Study 1	Costliness of empathy	318	3	954	-
Study 2	Eliciting peer praise	115	-	115	-
Study 3	Peer praise on empathy	328	15	4920	Praise-Empathy (n=1559), Praise-Describe (n=1801), Control (n=1560)
Study 4	Happiness as mediator	363	1	363	Praise-Empathy (n=127), Praise-Describe (n=13), Control (n=83)
Study 5A	Mediation analysis	338	20	6760	Praise-Empathy (n=1559), Praise-Describe (n=1801), Control (n=1560)
Study 5B	Mediation analysis	624	3	1872	Praise-Empathy (n=866), Praise-Describe (n=141), Control (n=865)
Total	-	2086	-	14984	-

Table A.1: Summarizing information on studies.

General handling of attrition For all studies, we evaluated attrition and its possible effects on our results in the same manner. We present for each study an attrition evaluation plot, whereby the x-axis presents in order questions posed to the respondents in the survey experiment. The y-axis denotes the proportion of respondents who attrited (compared to the original starting sample). We indicate through colored vertical lines where Pre-Treatment, Treatment, Mediator (or Other), and Outcome variables are measured. When large proportions of attrition occur at specific moments of the survey, it can become quickly clear to the researcher if these are at key points of the study – such as if it was treatment-induced attrition, which would most directly and problematically affect estimation of average treatment effects. Throughout our studies we see very low attrition (an average of 5%) with no obvious correlations with introduction of treatment.

B Further details on Measurements

B.1 Treatments

Measurement of peer praise We elicit naturalistic peer praise in Study 2 (see details on the Study in Appendix Section E) in the following manner:

1. We ask respondents to provide feedback on two tasks a real adult has performed – the FEEL and DESCRIBE tasks and explained what each task entailed and an example drawn image of a person.

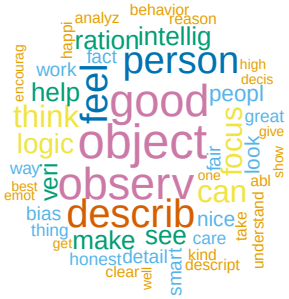
2. Respondents are asked to think of language that would admire or encourage the participant for choosing and doing the FEEL/DESCRIBE, especially positive things that can be said to people who choose to empathize/be objective to others in order to encourage them. Respondents then are asked for three words, then a full sentence. Finally respondents are asked to select how they feel about people who choose and engage in empathetic/objective behavior in a thermometer from 0-10 with zero as least warm and 10 as most warm.

3. To encourage respondents to think and write genuinely, we ask respondents in a series of follow up questions to tell us what the likelihood participants who are shown their

words will believe that they are genuine, and give respondents the opportunity to return and edit their responses if they desire.

We collected the words used by respondents to praise empathetic behavior and created a word cloud, with a short sentence above indicating the average feeling thermometer value for that behavior, calculated from Study 2 participants. This constitutes the main peer praise for empathy treatment, replicated here and found in the main text as well. We similarly create a peer praise for describe treatment for our robustness checks. Both are found in Figure B.1.

(a) Peer praise for empathy



(b) Peer praise for objectivity

Figure B.1: **Left panel (a) main peer praise for empathy treatment; right panel (b) peer praise for objectivity.**

B.2 Dependent variables

Measurement of main (choice) task Our main dependent variable is a forced-choice task selection between FEEL and DESCRIBE; for more detail please see Appendix Section C.1.

Measurement of reservation wage Our secondary dependent variable is an incentivized reservation wage elicitation for the FEEL task; for more detail please see Appendix Section C.1.

B.3 Mediators

Happy Study 5 immediately after the randomization of treatment, respondents were asked about their happiness *only* developed from an emotion scale by Harmon-Jones, Bastian and Harmon-Jones (2016). We specifically focus on the measurement of respondent emotion *in the moment*, so as to avoid conflating emotions across the experience of the overall survey with the emotions related to the treatment. Below is the phrasing of the

happiness measure:
`This scale consists of a number of words that describe feelings and emotions. Read each item and then mark the appropriate answer in the space next to the word. Indicate to what extent you feel this way RIGHT NOW.`
scale: very slightly or not at all/ a little/ moderately/ quite a bit/ extremely
emotions: happy/enjoyment/liking

B.4 Attention checks

Given concerns of greater online fatigue and inattentiveness during the COVID-19 global pandemic (see Peyton, Huber & Coppock, Working paper 2020), we follow Peyton et al.'s work, and the work of others on the usage of attention checks in online surveys (see for example Berinsky et al. (2019)), and incorporate two pre-treatment attention check questions for Studies 4, 5A and 5B. The first attention check is styled in a multiple choice and the second via a grid question (see Figures B.2-B.3) to capture most attentive respondents as well as least Berinsky et al. (2019). The multiple choice (`attentionMC`) and screening questions in the grid (`attentionG`) are drawn directly from Berinsky et al. (2019), while the filler questions in the grid are designed to elicit non-politically oriented opinions from respondents so as to minimize possible priming effects downstream.

B.5 Respondent covariates

Each study asked a series of respondent-level covariates within the surveys; for ease we present information on the collection and timing of each of respondent covariates across studies in Table B.2.

C Task descriptions

C.1 Main choice task

The main choice task that appears throughout Studies 1, 3, 4 and 5, entails a practice round, where respondents practice both FEEL and DESCRIBE activities. We describe the practice and main task below. For the practice, main task (and reservation wage task) images are drawn from the **Faces Data** in Chicago Faces and Harvard Faces Databases, randomized among the following features: Race=Black/White, Gender=Male (no variation), Valence=Angry/Fearful; images are randomized *without replacement* within respondent. See Figure C.8 for example draws of faces.

Practice Trial All respondents complete a practice trial after pretreatment covariates are collected and before the main randomization (praise).

We are interested in what sections people like to read in the newspaper. This might affect what they learn from articles and how they feel about the issues discussed in them. We also want to see if people are reading the questions carefully. To show that you've read this much, please mark both the classified and none of the above options below. That's right, just select these two options only.

Regardless of how frequently you read the newspaper, what would you say are your favorite newspaper sections to read? (please check all that apply)

- National
- Local
- Real estate
- Comics
- Classified
- Style
- Sports
- Business
- Science and technology
- Opinion
- None of the above
- All of the above

Figure B.2: MC attention check

In the following, you will complete a task. You will first complete a practice trial, which will help you become familiar with the task. After the practice trial.
On the trial, you will see two decks of shuffled cards: the deck on the left will always be labeled DESCRIBE and the deck on the right will always be labeled FEEL. You should choose between these

In the grid below, you will see a series of statements. Please tell us whether you agree or disagree with each statement.

	Agree strongly	Agree	Neither agree nor disagree	Disagree	Disagree strongly
The best sport to watch live is baseball	○	○	○	○	○
Facebook is the best social media platform	○	○	○	○	○
Two is greater than one	○	○	○	○	○
Football is more interesting than basketball	○	○	○	○	○
Please click the "neither agree nor disagree" response	○	○	○	○	○
Twitter is more engaging than Instagram	○	○	○	○	○
Soccer is more fun to play than hockey	○	○	○	○	○

Figure B.3: Grid attention check

	Study 1	**Study 2**	**Study 3**	**Study 4**	**Study 5 (A & B)**
State of residence	Post DV	-	Post T/DV	Pre T	Pre T
Age	Post DV	Post DV	Post T/DV	Pre T	Pre T
Sex	Post DV	Post DV	Post T/DV	Pre T	Pre T
Education	Post DV	Post DV	Post T/DV	Pre T	Pre T
Race	Post DV	Post DV	Post T/DV	Post T/DV	Pre T
Income	Post DV	Post DV	Post T/DV	Post T/DV	Pre T
Religion	Post DV	Post DV	Post T/DV	Post T/DV	Pre T
Party	Post DV	Post DV	Post T/DV	Post T/DV	Pre T
Ideology	Post DV	Post DV	Post T/DV	Post T/DV	Post T/DV
Trump approval	Post DV	-	Post T/DV	Post T/DV	Post T/DV
Biden approval	-	-	-	-	Post T/DV
Baseline empathy	Post DV	-	Post T/DV	Post T/DV	Post T/DV

Table B.2: Measurement of respondent covariates across studies. T indicates when treatment (peer praise) was measured, DV indicates when dependent variables are measured. In Studies 1 and 2 no treatments were manipulated.

decks. Once you choose a deck, you will then see an image of a person. The decks include the same images. Depending on which deck you have chosen, you will be given one of two possible sets

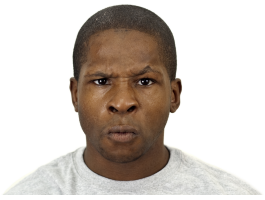


Figure C.4: Race=Black, Valence=Angry

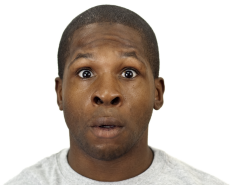


Figure C.5: Race=Black, Valence=Fearful

Figure C.6: Race=White, Valence=Angry

Figure C.7: Race=White, Valence=Fearful

Figure C.8: **Example faces from Chicago Faces Database.**

of instructions.
If you choose from the deck labeled DESCRIBE, you will be told to be objective and focus on the external features and appearances of the person in the image. When completing this kind of trial, try to be as objective as possible. To be objective, do not let yourself get caught up in imagining what this person feels. On these trials, describe the **age**, **gender** and **race** of the person.
If you choose from the deck labeled FEEL, you will be told to have empathy and focus on the internal feelings and experiences of the person in the image. When completing this kind of trial, try to feel as much empathy as possible. To be empathetic, let yourself get caught up in imagining what this person feels. On these trials, describe the **feelings** and **experiences** of the person.
You are free to choose from either deck on any trial, and should feel free to move from one deck to the other whenever you choose. If one deck begins to seem preferable, feel free to choose that deck more often. Overall, this task will take the same amount of time regardless of which deck you choose.

Now you will complete a practice trial of the task; later on, you will turn over to the task.
Please click on one of the decks.

[Depending on what respondent clicks first, rotate questions below]

[Deck chosen is highlighted; image is presented.]

FEEL

1. Please write a sentence describing the **feelings** and **experiences**

of this person. [Open-ended sentence. Cannot proceed without writing, min time 10 seconds.]

2. Please write three words that describe the **feelings** and **experiences** of this person. [Three open-ended slots. Cannot proceed without writing in all slots, min time 10 seconds.]

3. How do you feel about this person? [Feeling thermometer]

Please click on the other deck. [Forced choice]
[Other deck is highlighted; image is presented]

DESCRIBE

1. Please write a sentence describing the **age , gender** and **race** of this person. [Open-ended sentence. Cannot proceed without writing, min time 10 seconds.]

2. Please write three words that describe the **age , gender** and **race** of this person. [Three open-ended slots. Cannot proceed without writing in all slots, min time 10 seconds.]

3. How do you feel about this person? [Feeling thermometer]

Main task After the practice round, respondents enter into the main task (with or without treatment praise, depending on the Study).

In the following trial, you will repeat the task you did in the practice round, where you will be asked to choose the deck you prefer each time, presented with a person in an image, and then asked to answer questions related to the deck you chose.
Recall: You are free to choose from either deck.

1. [Present two labeled decks; do not allow any clicking until after the treatment is drawn.]

2. Then print on the page: Please click on one of the decks.

3. [Deck chosen is highlighted; image is presented.]

4. [Provide trial questions based on whichever deck is selected by the respondent.]

C.2 Reservation wage task description

In the next task, you will be making choices between real amounts of money. You will see several choices to make between the two decks of cards, exactly like the tasks you completed earlier.

For each choice between the decks, the DESCRIBE deck asks you to be objective and write about the age and race of a person, and the FEEL deck asks you to be empathetic and write about the internal experiences and feelings of a person. In all cases, the persons shown faces similar to those you saw earlier in the experiment. This time, you will see a real payment for completing a trial from each deck, for each choice. Please select the option that you prefer for each of the choices. There are no accurate or inaccurate answers. A random draw from one of the sets of choices will be enacted, and you will be directed to the deck you chose under that choice set, and paid the amount for that choice. These choices are thus **real decisions with real pay .**

[Present list of paired options of decks for respondent to click on.]

Wages for DESCRIBE are always $1.00; wages for FEEL range from $0.99 to $1.13 in 1 cent increments. Each time an option for a pair is clicked on, the respondent will see the sentence below the pair "I would prefer to conduct task DESCRIBE/FEEL for Y amount, over task FEEL/DESCRIBE for Z amount.'

Recall your preferred choices for wages for DESCRIBE and FEEL.

[Randomly assign one of the paired options in the **Real Wage Task** to execute. Highlight the selected row.]

A random draw of the paired choices you have made has been selected: you will now conduct task X for Y amount. Your Y pay will be added to your survey pay at the end of this survey.

D Study 1: Costs of Empathy

Study 1 was fielded in September 2020, with a total of 318 respondents. The purpose of the study was to establish the baseline costliness of empathy. Figure D.9 presents the consort diagram for Study 1. Descriptive statistics on respondent covariates are presented in Table D.3. Respondents were also asked about their beliefs on how often other respondents on the platform chose the FEEL andDESCRIBE tasks, and what they thout others' beliefs about empathy and objectivity were (see Table D.4). Other than measuring respondents' behavioral choices to establish baseline costs of empathy, we also directly asked respondents to rate the FEEL andDESCRIBE tasks for difficulty using the NASA task load; summary statistics for answers to these questions are presented in Table D.5 and differences between the answers by task type are in Table D.6.

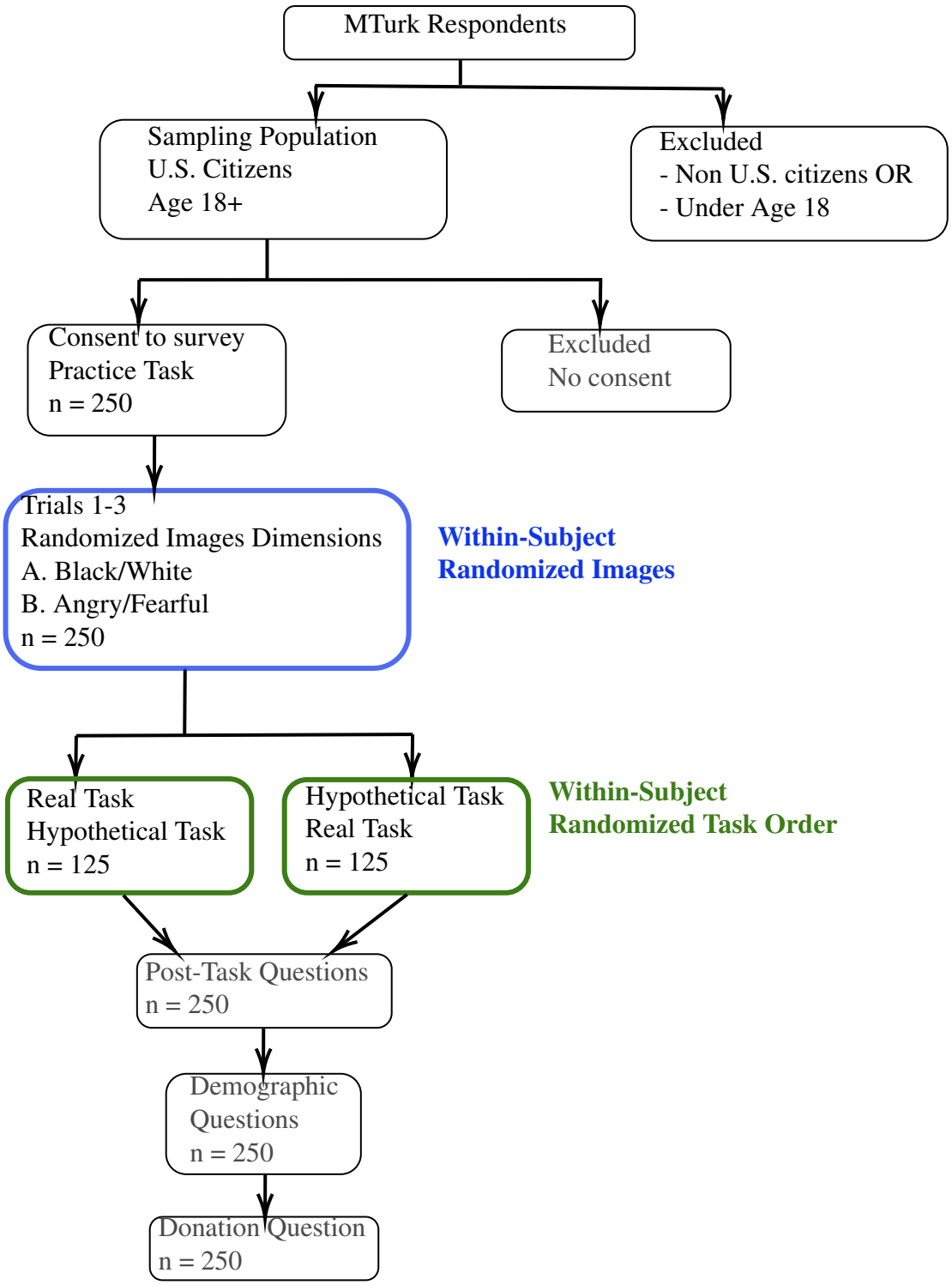


Figure D.9: **Study 1: Consort Diagram**

	Level	N	%
Sex	Female	97	30.5
	Male	155	48.7
	Missing	66	20.8
Race	White	195	61.3
	Asian	1	0.3
	Black or African American	33	10.4
	Native Hawaiian or Pacific Islander	13	4.1
	Other	7	2.2
	Missing	69	21.7
Education	Associate degree	17	5.3
	Bachelor's degree (BA/BS)	149	46.9
	High school or equivalent (GED)	14	4.4
	Kindergarten to 8th grade	1	0.3
	Master's degree (MA/MS/MBA)	43	13.5
	Medical (MD), law (JD) or other doctoral degree (PhD)	2	0.6
	No schooling completed	1	0.3
	Some college, but did not complete a degree	25	7.9
	Missing	66	20.8
Income	100k or more	14	4.4
	25k to less than 50k	91	28.6
	50k to less than 75k	91	28.6
	75k to less than 100k	29	9.1
	Less than 25k	27	8.5
	Missing	66	20.8
Religion	Atheist/agnostic	45	14.2
	Buddhist	9	2.8
	Hindu	4	1.3
	Jewish	6	1.9
	Muslim	7	2.2
	Nothing in particular	24	7.5
	Orthodox (Greek or Russian)	1	0.3
	Protestant	53	16.7
	Roman Catholic	103	32.4
	Missing	66	20.8
Party	Democrat	62	19.5
	Independent	33	10.4
	Lean Democrat	21	6.6
	Lean Republican	13	4.1
	Republican	63	19.8
	Strong Democrat	34	10.7
	Strong Republican	26	8.2
	Missing	66	20.8
Ideology	Conservative	34	10.7
	Liberal	75	23.6
	Moderate	37	11.6
	Slightly conservative	24	7.5
	Slightly liberal	23	7.2
	Very conservative	25	7.9
	Very liberal	34	10.7
	Missing	66	20.8
Age	+70	2	0.6
	20-30	98	30.8
	31-40	93	29.2
	41-50	20	6.3
	51-60	27	8.5
	61-70	10	3.1

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Belief people choose Feel task	252	57.421	23.959	3.000	39.750	79.000	100.000
Belief people choose Describe task	250	67.320	17.788	2.000	56.250	80.750	100.000
Belief people think empathy is good	250	72.396	18.783	1.000	58.500	86.000	100.000
Belief people think objectivity is good	250	72.128	16.948	1.000	63.000	85.000	100.000

Table D.4: Descriptive Statistics - Empathy norms

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Describe task mentally demanding	252	2.702	1.232	1.000	2.000	4.000	5.000
Feel task mentally demanding	252	2.937	1.043	1.000	2.000	4.000	5.000
Describe task hard to accomplish	252	2.385	1.037	1.000	2.000	3.000	5.000
Feel task hard to accomplish	252	2.762	1.085	1.000	2.000	3.000	5.000
Describe task raised insecurity	252	2.095	1.177	1.000	1.000	3.000	5.000
Feel task raised insecurity	252	2.329	1.170	1.000	1.000	3.000	5.000
Describe task done successfuly	252	3.762	1.005	1.000	3.000	4.000	5.000
Feel task done successfuly	252	3.575	1.048	1.000	3.000	4.000	5.000

Table D.5: Descriptive Statistics - NASA task load

Task	**Demanding**	**Hard**	**Insecure**	**Successful**
Objective (DESCRIBE)	2.702	2.385	2.095	3.762
Empathy (FEEL)	2.937	2.762	2.329	3.575
Difference	0.234 (p=0.0217)	0.377 (p=1e-04)	0.234 (p=0.0256)	-0.187 (p=0.042)

Table D.6: Task load summary. Mean values reported (choices from 1-5).

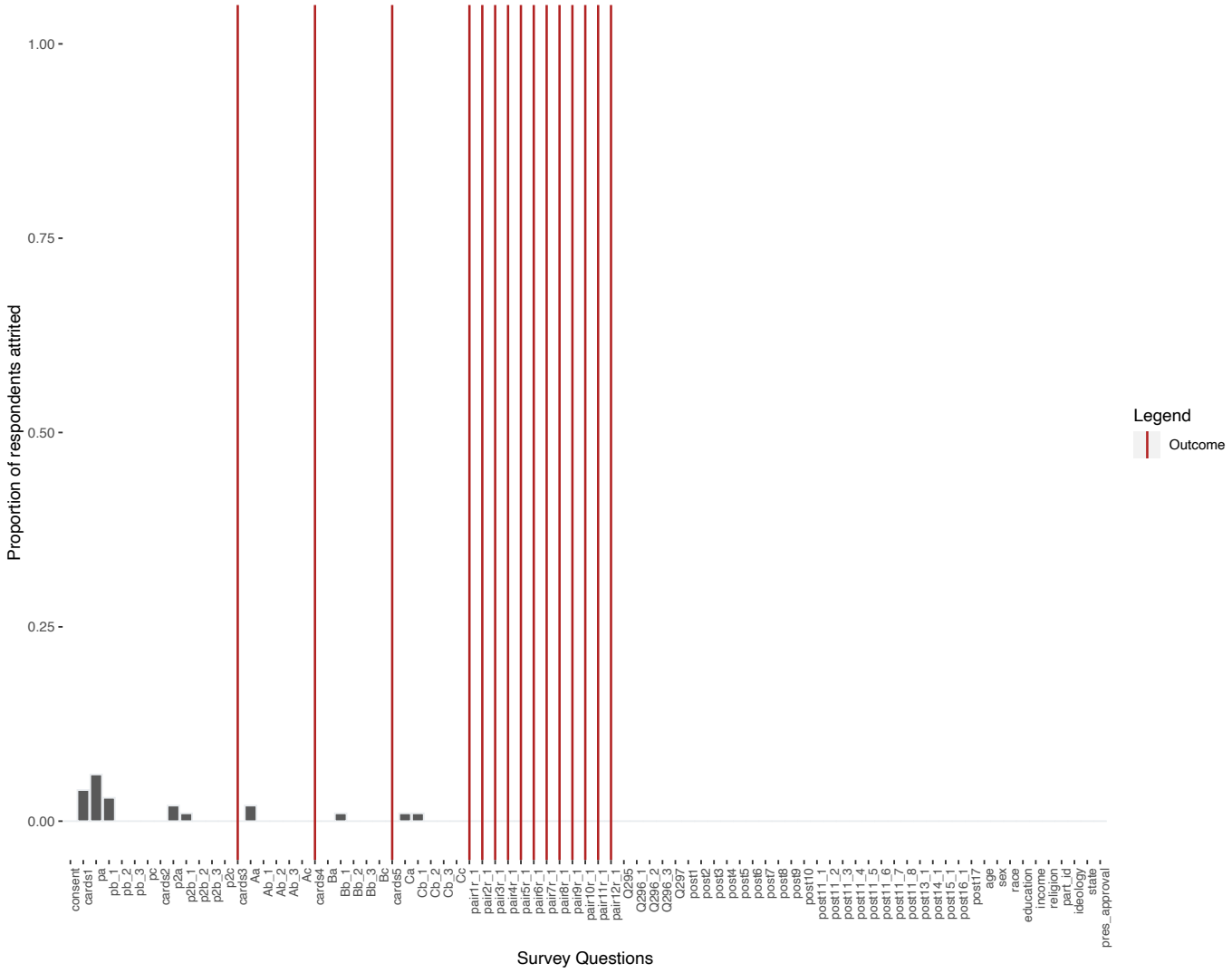


Figure D.10: **Attrition across survey questions.**

Study 1 Attrition

E Study 2: Eliciting Naturalistic Praise (non-experimental)

Study 2 was fielded in September 2020 with a total of 115 respondents. The purpose of the study was to elicit naturalistic peer praise for the empathy and objective tasks. The consort diagram for Study 2 is presented in Figure E.11. Table E.7 presents respondent covariate descriptives from the study. No attrition occurred in the study. Respondents were asked to write sentences and words that would praise peers who engaged in empathetic/objective behavior due to doing the FEEL and DESCRIBE tasks. Figure **??** presents a plot of the words that occur differentially across the words elicited for praising FEEL and DESCRIBE.

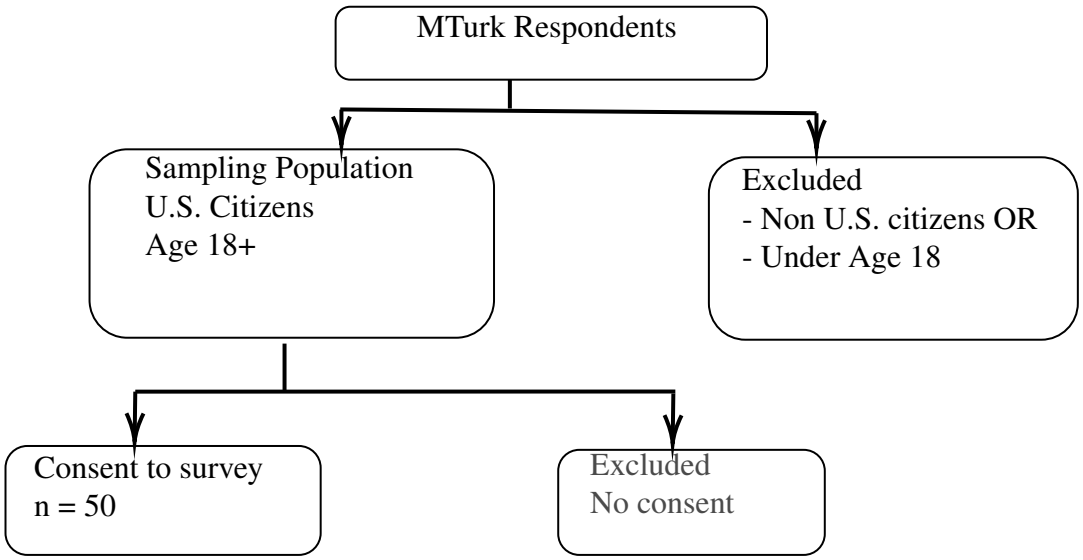


Figure E.11: Study 2: Consort Diagram

	Level	N	%
Sex	Female	39	33.9
	Male	76	66.1
Race	Asian	2	1.7
	Black or African American	15	13.0
	Native Hawaiian or Pacific Islander	9	7.8
	Other	8	7.0
	White	81	70.4
Education	Associate degree	11	9.6
	Bachelor's degree (BA/BS)	53	46.1
	High school or equivalent (GED)	9	7.8
	Master's degree (MA/MS/MBA)	22	19.1
	Medical (MD), law (JD) or other doctoral degree (PhD)	2	1.7
	Some college, but did not complete a degree	18	15.7
Income	100k or more	9	7.8
	25k to less than 50k	32	27.8
	50k to less than 75k	39	33.9
	75k to less than 100k	18	15.7
	Less than 25k	17	14.8
Religion	Atheist/agnostic	37	32.2
	Buddhist	1	0.9
	Jewish	1	0.9
	Mormon	1	0.9
	Nothing in particular	17	14.8
	Orthodox (Greek or Russian)	1	0.9
	Protestant	22	19.1
	Roman Catholic	35	30.4
Party	Democrat	38	33.0
	Independent	17	14.8
	Lean Democrat	9	7.8
	Lean Republican	7	6.1
	Republican	23	20.0
	Strong Democrat	11	9.6
	Strong Republican	10	8.7
Ideology	Conservative	19	16.5
	Liberal	36	31.3
	Moderate	20	17.4
	Slightly conservative	8	7.0
	Slightly liberal	9	7.8
	Very conservative	8	7.0
	Very liberal	15	13.0
Age	+70	1	0.9
	20-30	51	44.3
	31-40	41	35.7
	41-50	9	7.8
	51-60	9	7.8
	61-70	3	2.6
	Missing	1	0.9

Table E.7: **Study 2 Respondents**. Total number of respondents 115.

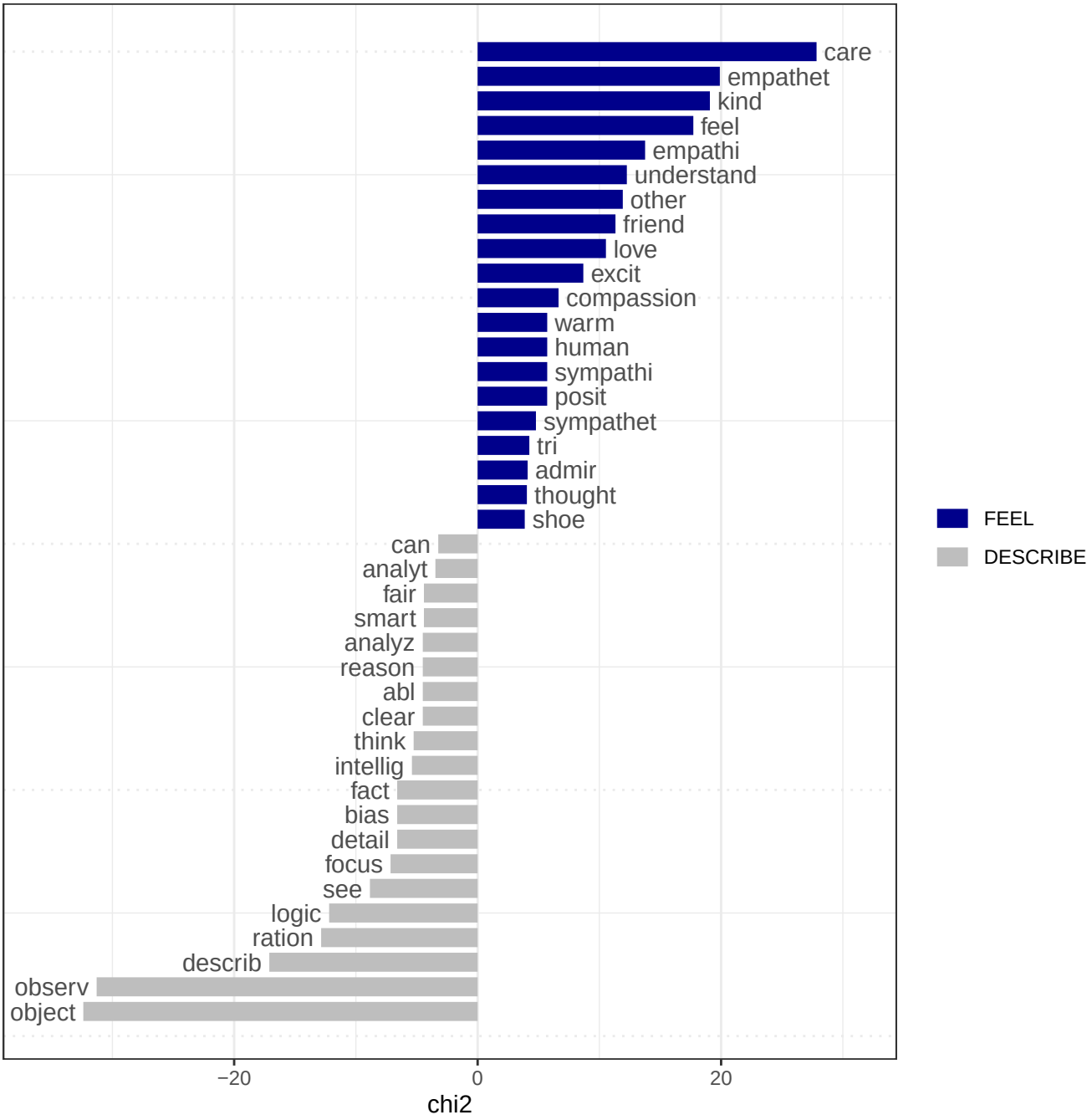


Figure E.12: **Keyness plot of words for empathy (FEEL) versus objective (DESCRIBE) tasks.** Figure plots the results of a keyword of features comparing their differential associations with providing language in praise of peers who engage in empathy (FEEL) versus objective (DESCRIBE) tasks, after calculating "keyness", a score for features that occur differentially across different categories. Here text for (FEEL) and (DESCRIBE) are the different categories.

As a further check on whether positive feelings are held towards people who exhibit empathetic or observational behaviors, we asked respondents to provide thermometer ratings towards people who exhibited these types of behaviors. We calculate the positive and negative sentiments for praise texts respondents generated for people who display empathetic

and objective behaviors respectively, using the Lexicoder Sentiment Dictionary and verify whether the thermometer ratings are positively correlated with positive text sentiments and negatively correlated with negative text sentiments. Figures E.13 and E.14 present linear association results that suggest the same.

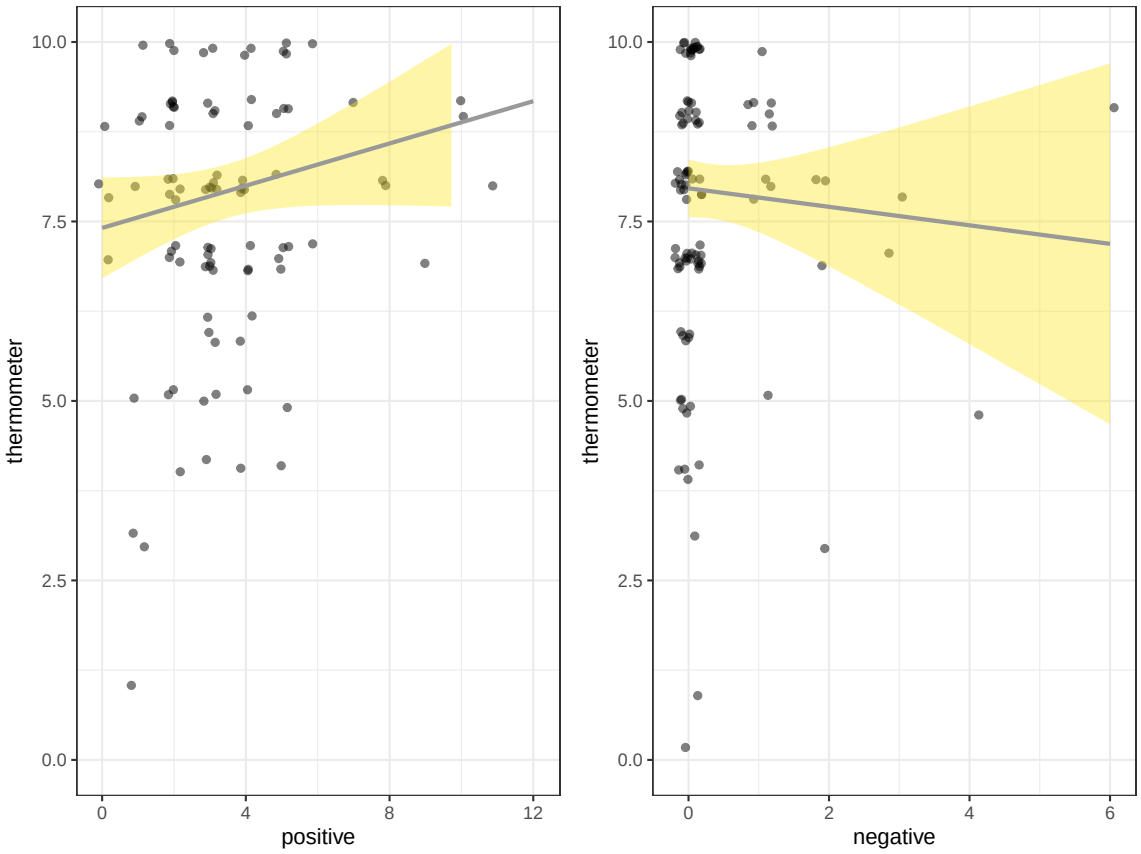


Figure E.13: Correlation between positive and negative text sentiments for generated texts of praise for empathetic behavior with thermometer ratings for people who engage in empathetic behavior.

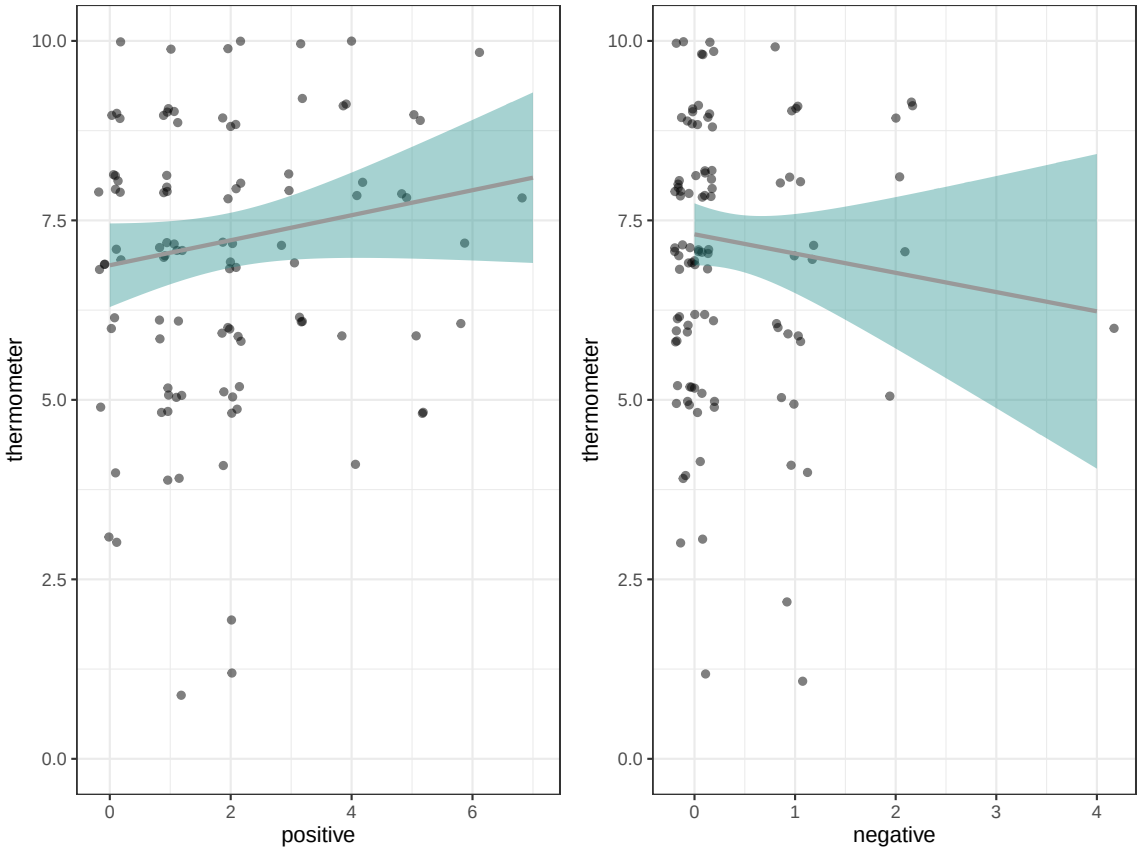


Figure E.14: Correlation between positive and negative text sentiments for generated texts of praise for objective behavior with thermometer ratings for people who engage in objective behavior.

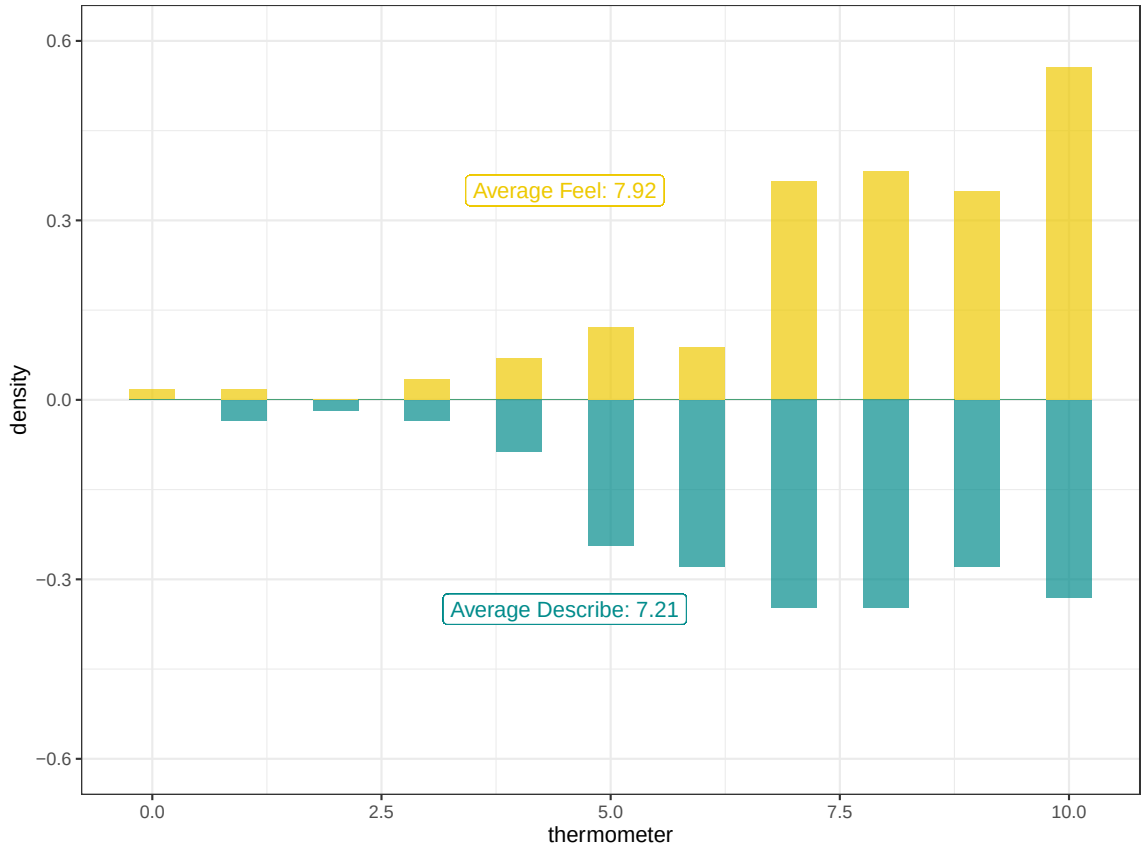


Figure E.15: Distributions of thermometer ratings towards peers who exhibit empathetic behavior (top) and towards peers who exhibit objective behavior (bottom).

Study 2 Attrition

F Study 3: Praise Lowers the Cost of Empathy

Study 3 was fielded in November 2020 with 328 respondents. The primary purpose of the study was to establish whether peer praise (for empathy) could encourage empathetic behavior. We randomized peer praise for empathetic behavior, peer praise for objective behavior (taken from Study 2) and a control arm of no intervention and measured respondents' choice of task between FEEL and DESCRIBE. Secondarily, we were interested in evaluating whether peer praise might change reservation wages for the FEEL task. Figure F.16 depicts the consort diagram for Study 3.

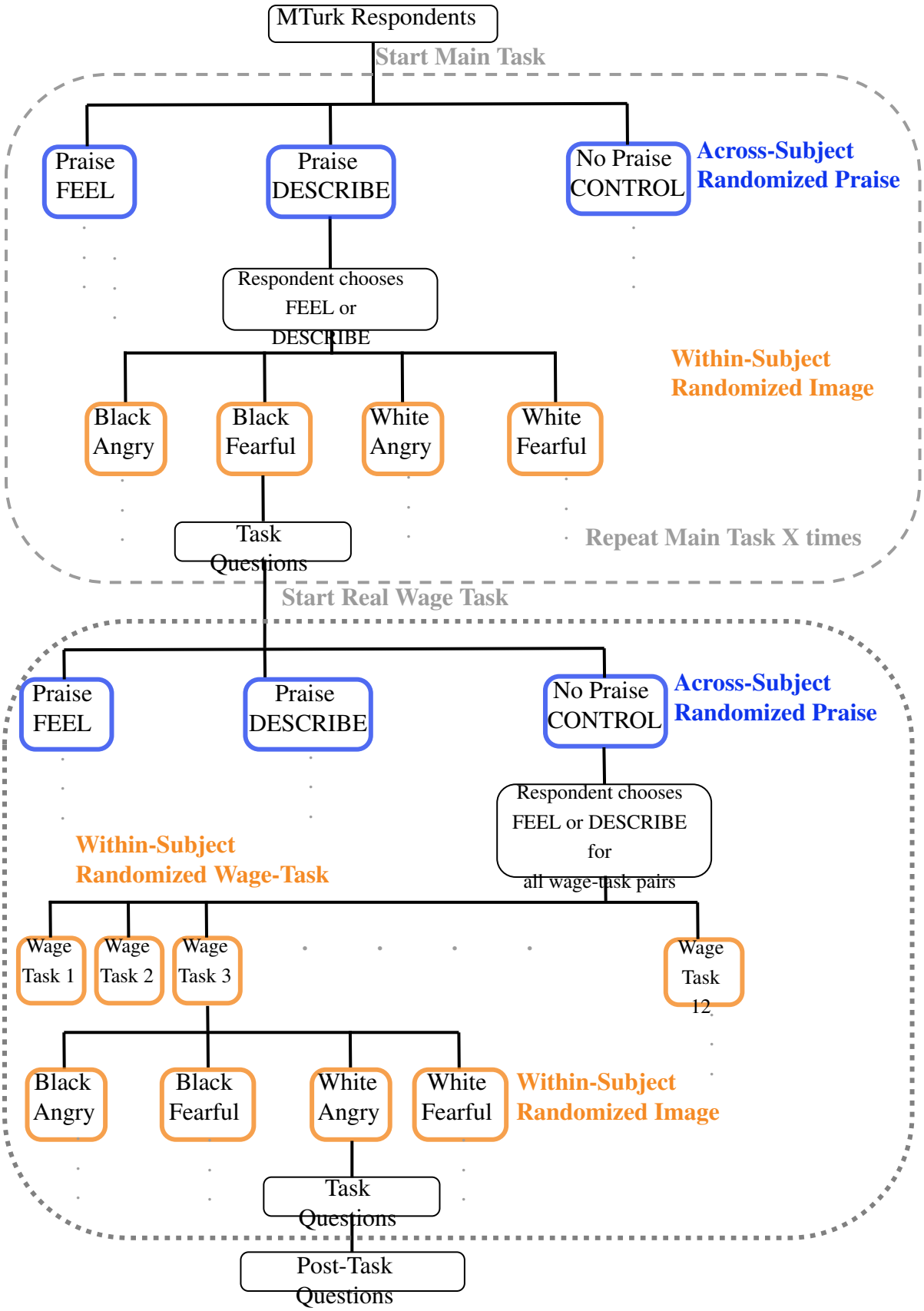


Figure F.16: **Study 3 Consort Diagram**

	Log Odds	95% CI	Odds Ratio	95% CI
Intercept	-0.442	[-0.625,-0.259]	0.643	[0.535,0.772]
Peer praise for empathy	0.182	[0.025,0.339]	1.200	[1.025,1.404]

Table F.8: Peer praise effect on task choice.

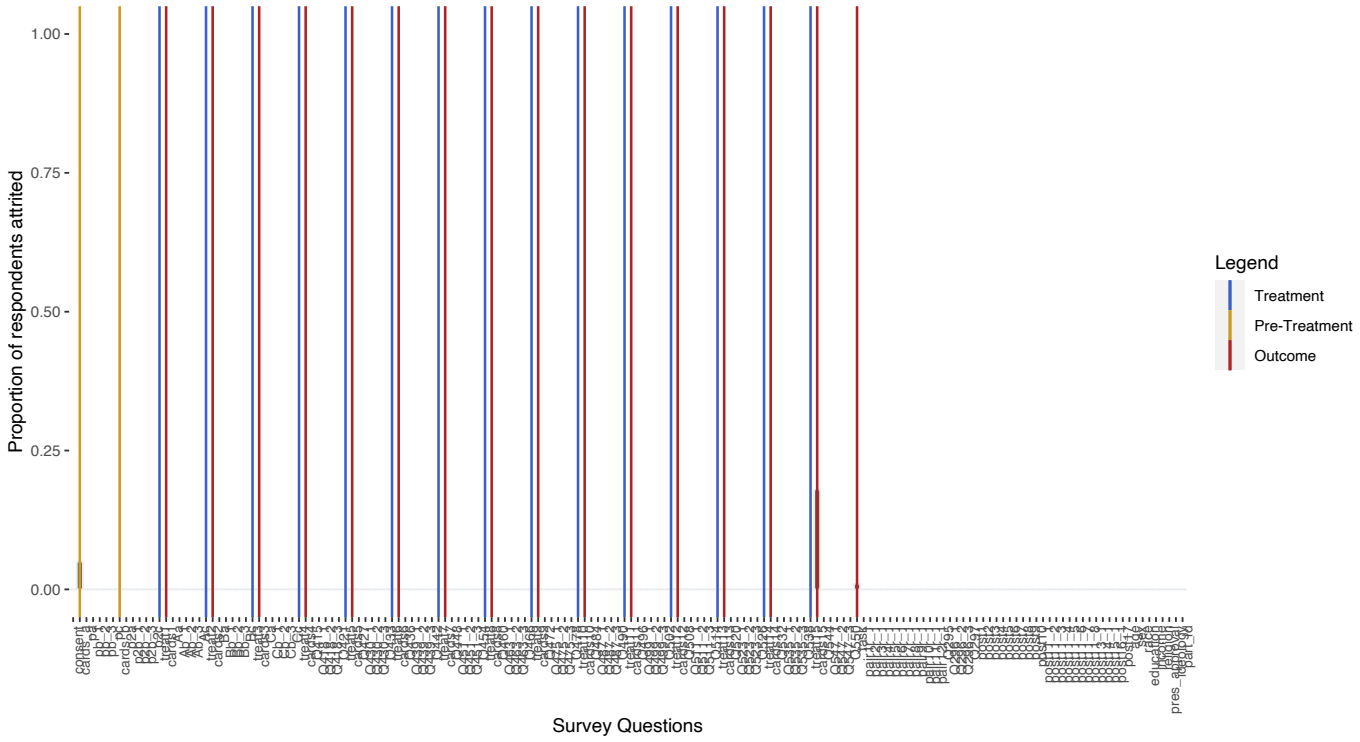


Figure F.17: **Attrition across survey questions:**

Study 3 Attrition A total of 75 attrited from the survey. Of those,18.7% attrited during the first set of instructions, 34.7% attrited during the practice round, and 21.3% attrited during the post task questions. Attrition is not associated with praise treatment, or randomization of images. Respondents who were primed with praise FEEL, were 0.5% less likely to attrite (baseline is 0.01) than compared to respondents who received the Control (no praise). This finding is not statistically significant ($p = 0.1$). Respondents who saw an image with a black person, were 1.8% less likely to attrite (baseline is 0.55) than compared to respondents who received an image with a white person. This finding is not statistically significant ($p = 0.4$). Respondents who saw an image with an angry person, were 0.4% less likely to attrite (baseline is 0.517) than compared to respondents who received an image with a fearful person. This finding is not statistically significant ($p = 0.8$).

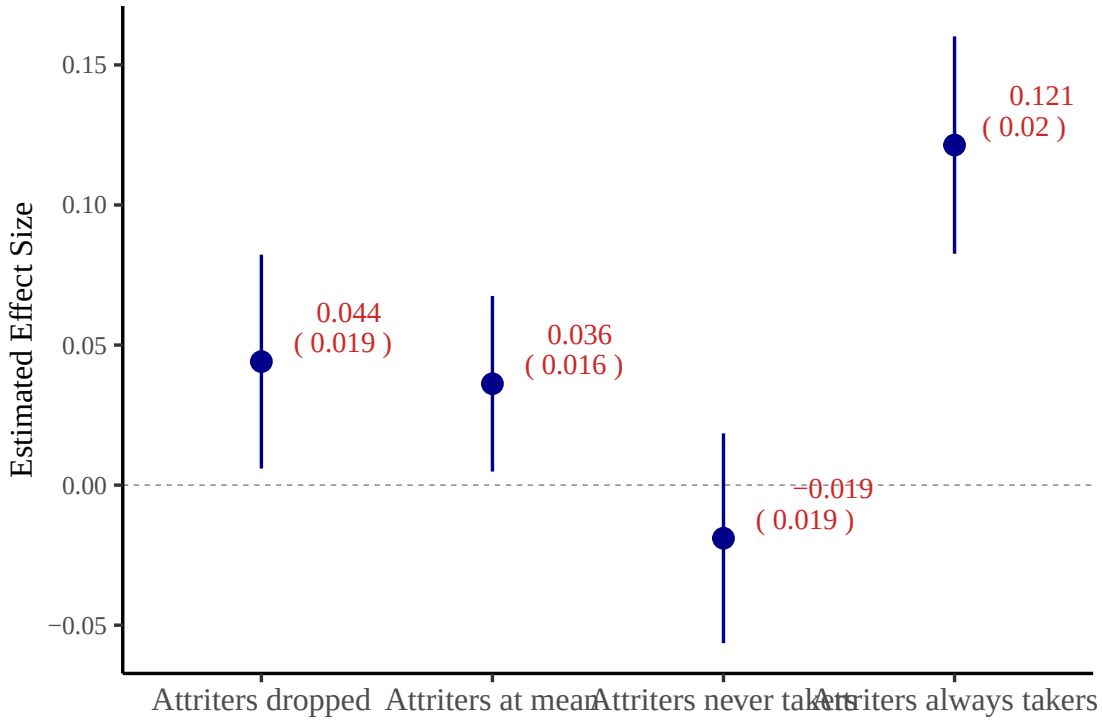


Figure F.18: **Missingness Imputed:** In all models, ‘describeORfeel‘ is regressed over ‘PraiseEmpathy‘, estimating the effect of praise FEEL on selecting to perform task FEEL. Attriters dropped is the main model presented in the paper, in which attriters are dropped from the analysis. Attriters at mean is coded such that attriters receive the mean value for ‘describeORfeel‘. Attriters never takers is coded such that attriters never take the praise treatment. Thus, ‘describeORfeel‘ receives the value 0 (DESCRIBE) when attrited respondents are treated with praise FEEL. Otherwise, attrited respondents receive a value selected from a distribution around the mean of controlled respondents.Attriters always takers is coded such that attriters always take the praise treatment. Thus, ‘describeORfeel‘ receives the value 1 (FEEL) when attrited respondents are treated with praise FEEL. Otherwise, attrited respondents receive a value selected from a distribution around the mean of controlled respondents.

We comparing peer praise for objective behavior vs peer praise for empathetic behavior on the willingness for respondents to choose the FEEL task in Figure **??**; in Figure **??** we check if peer praise for empathetic behavior can reduce the reservation wage for the FEEL task.

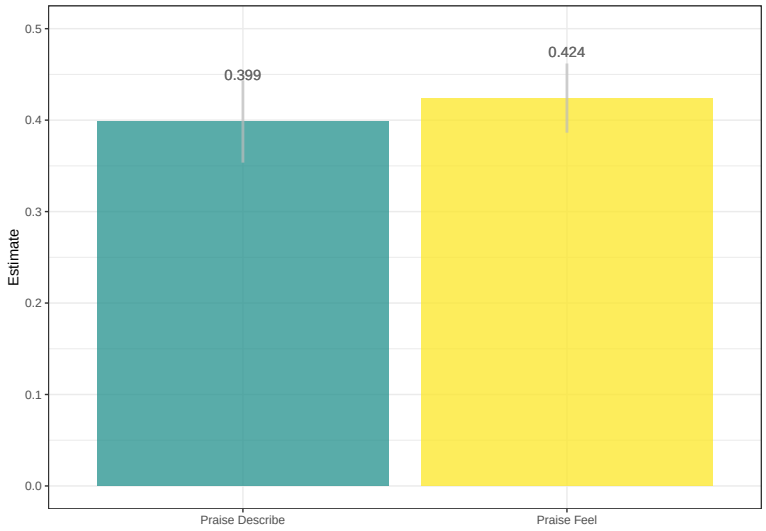


Figure F.19: **Effects of Praise for objective behavior on choosing FEEL over DESCRIBE task compared with the Praise for empathetic behavior.** Standard errors clustered at the respondent level. Difference in means is 0.026 with standard error 0.019.

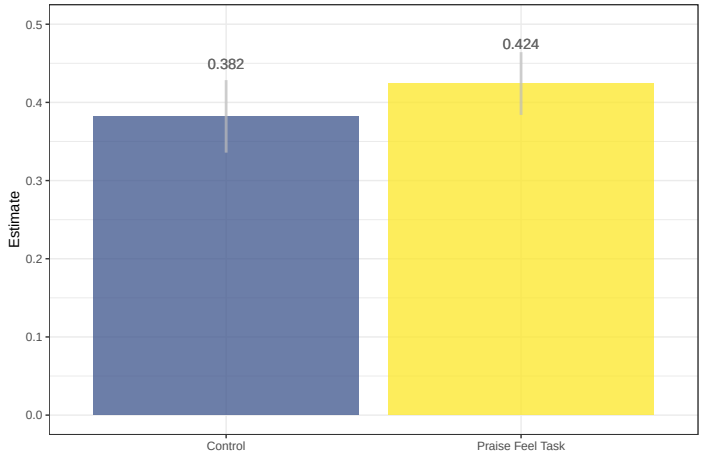


Figure F.20: **Effects of Praise for empathetic behavior on the reservation wage for the FEEL task compared with Control arm.** Difference in means is -0.03 with standard error 0.027.

Study 3 Robustness checks It could be that respondents who are peer praised into selecting the empathy task are in fact simply doing a less good job (so responding to the peer praise and then selecting to do less work afterwards). Here, we conduct a few empirical (observational) tests to try to see if respondents are indeed taking "short cuts". In our first test, we check if respondents who chose the empathy task under the peer praise treatment are similarly likely in using diverse (unique) words ("Unique tokens") compared with respondents who chose the empathy task under the control treatment. In our second test, we consider if respondents who chose the empathy task under treatment are more likely to use words from the peer praise wordcloud ("Proportion of wordcloud") – which would suggest short-cutting as well through simply applying words presented. In our last test, we check if the sentiment of words written in the empathy task is similar in the peer-praise group compared to the control group; if the former set of words are less positive, then it might suggest that respondents are actually not actually conducting the empathy task in the same manner. Our findings from the three tests are presented in Table F.9 and suggest that there does not seem to be evidence of shortcutting.

	DV: Unique tokens			DV: Proportion of wordcloud			DV: Text sentiment		
	Estimate	**s.e.**	**p**	**Estimate**	**s.e.**	**p**	**Estimate**	**s.e.**	**p**
Intercept	5.438	0.165	7.56e-237	0.164	0.013	1.03e-35	-0.2	0.028	7.96e-13
Peer praise	0.102	0.142	0.473	0.009	0.013	0.48	-0.016	0.032	0.614

Table F.9: Testing for shortcutting.

G Study 4: Peer praise increases reported happiness

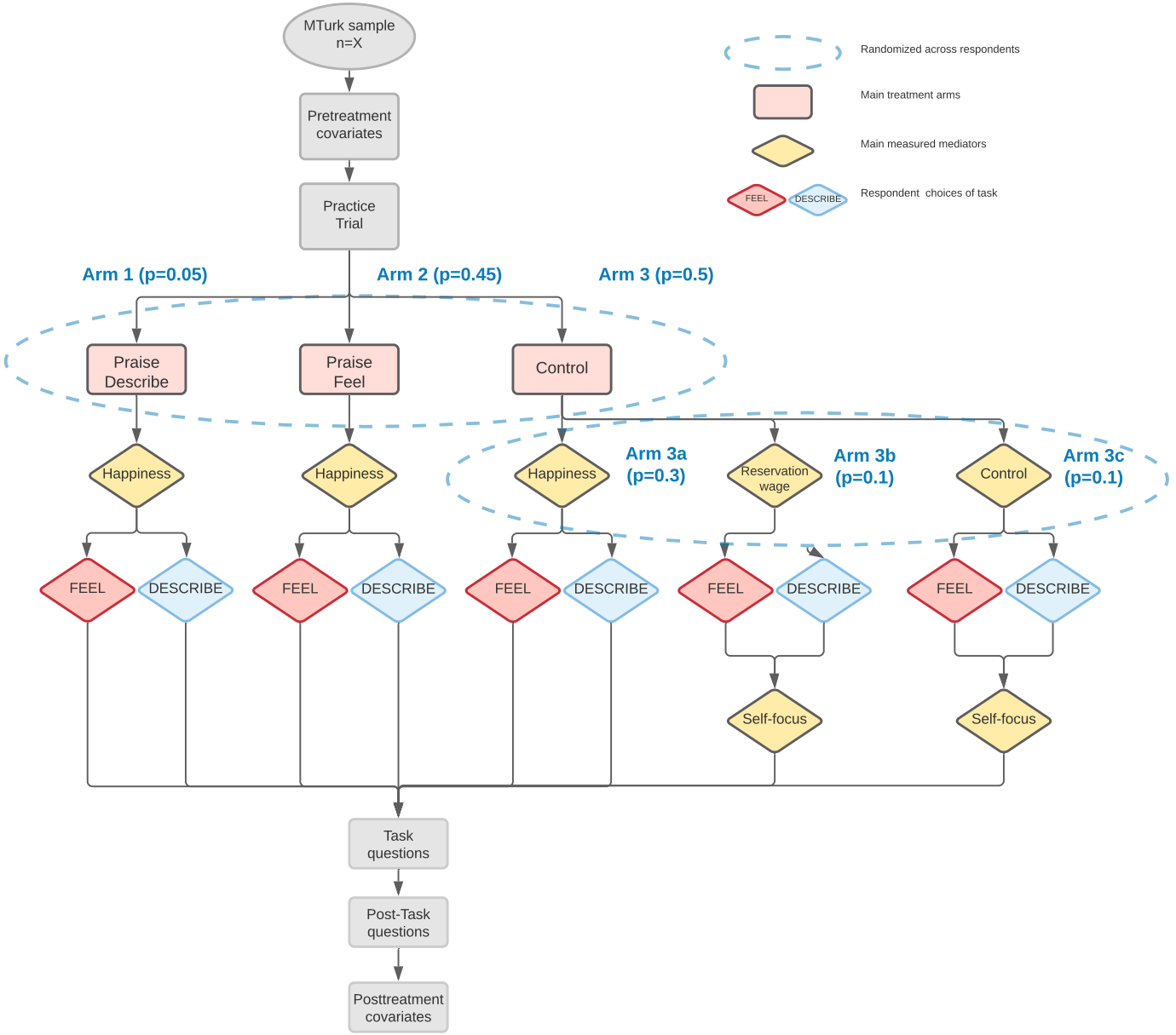


Figure G.21: **Study 4 Consort diagram**. Main arms labeled with probability of assignment in parentheses (probability out of total assignment).

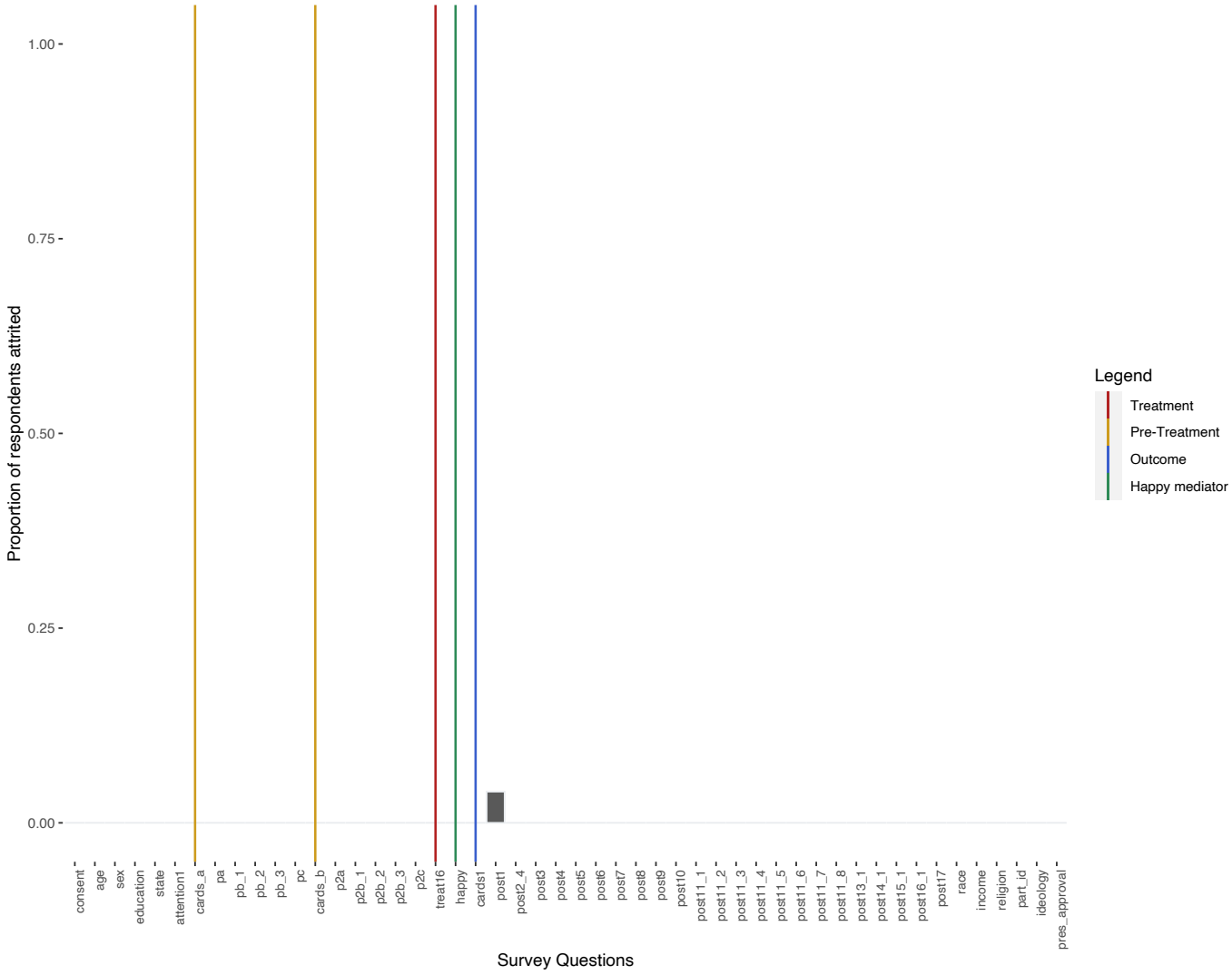


Figure G.22: **Attrition across survey questions.**

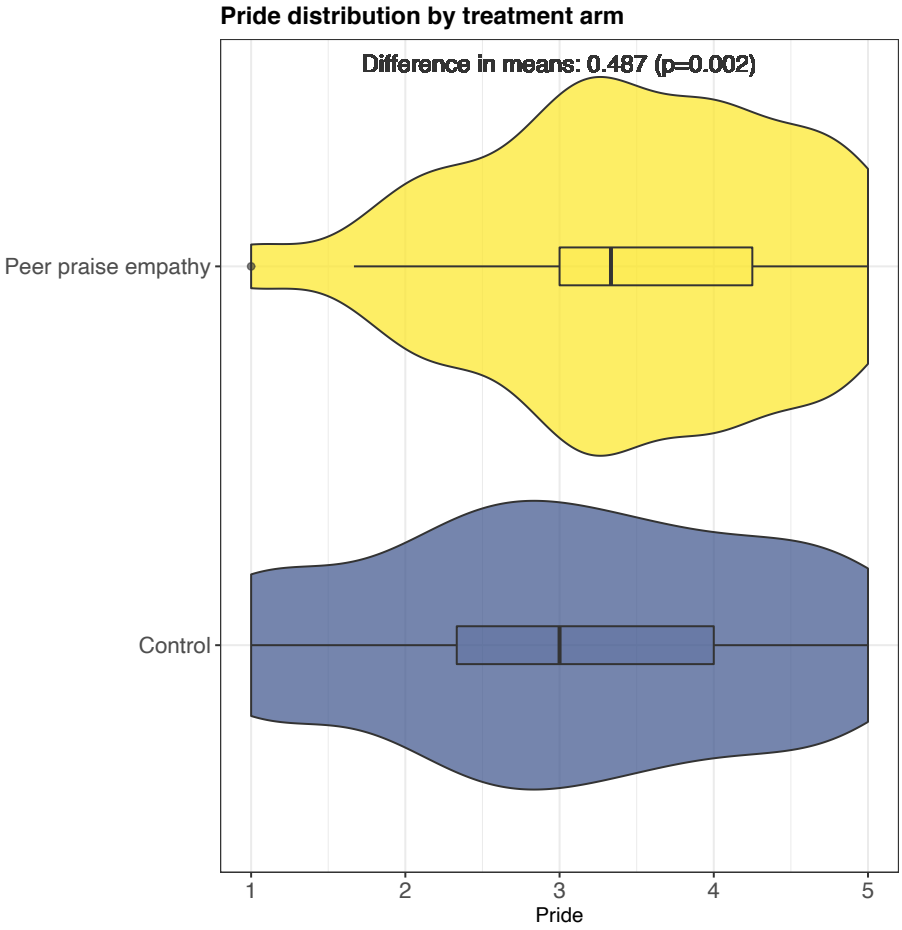


Figure G.23: Pride of respondents in peer praise (for empathy) and control groups.

Study 4 Attrition

H Study 5: Peer praise increases likelihood of empathy task through increased happiness

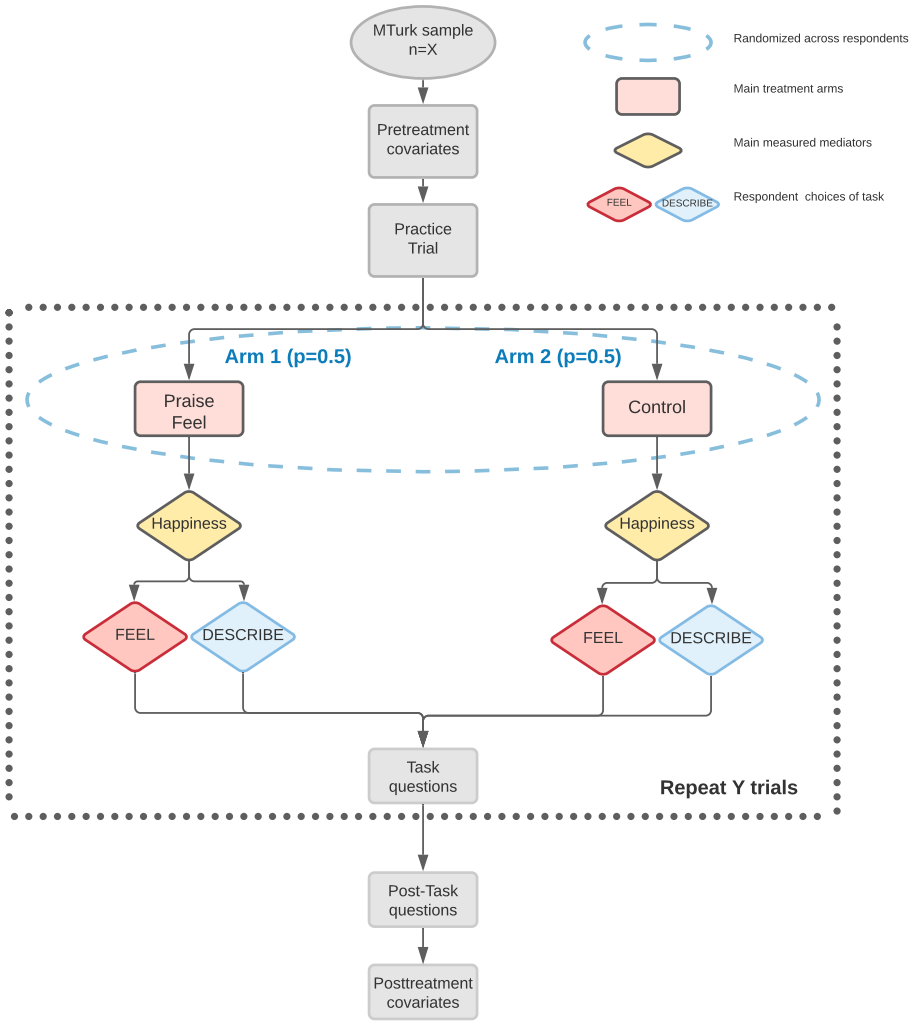


Figure H.24: **Study 5 Consort diagram**. Main arms labeled with probability of assignment in parentheses (probability out of total assignment). Dotted gray space encapsulates the main task, which is repeated for Y trials for each respondent, where for Study 5A Y is 20, while for Study 5B Y is 3.

Study 5 Attrition Attrition evaluation plots for 5A and 5B are presented in Figures H.25 and H.26 respectively.

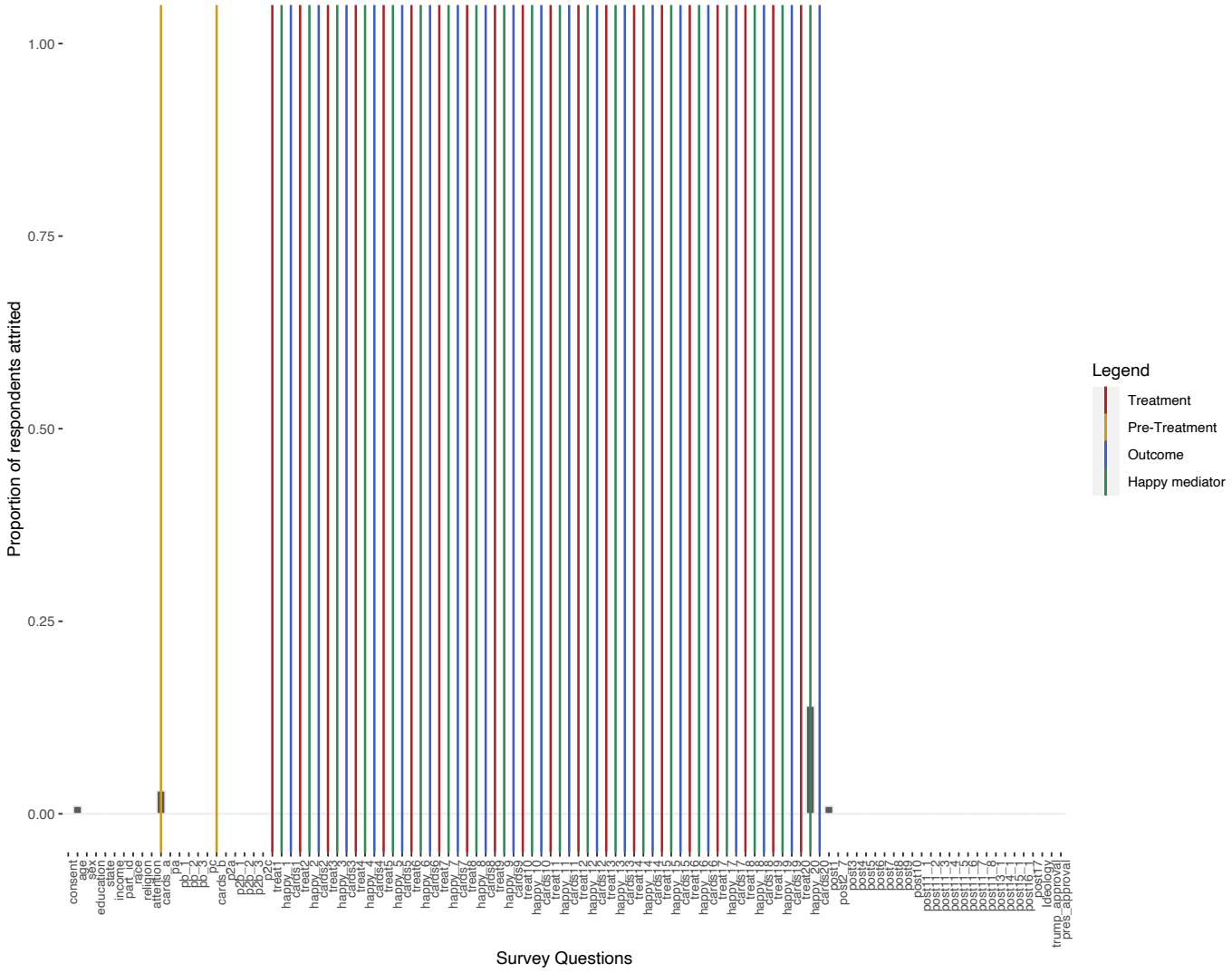


Figure H.25: **Study 5A: attrition across survey questions.**

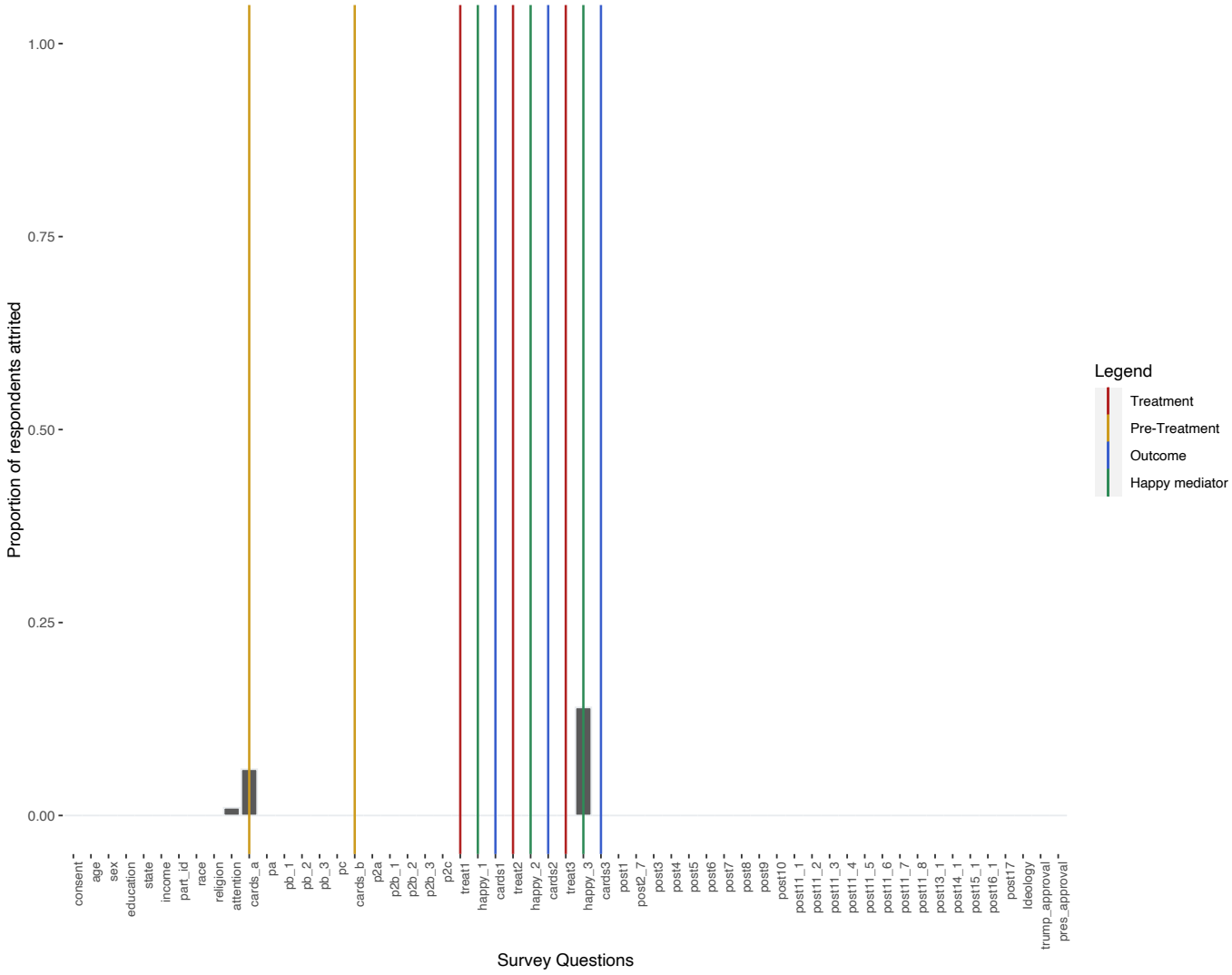


Figure H.26: **Study 5B: attrition across survey questions.**

Sensitivity analysis of mediation We analyze the mediating effect of happiness on the choice task variable using Imai, Keele and Yamamoto (2010) approach for model-based causal mediation analysis; the key assumption required is sequential ignorability. Thus we focus on the sensitivity parameter $\rho \equiv \text{Corr}(\epsilon_{i2}, \epsilon_{i3})$; sequential ignorability implies $\rho = 0$. We set ρ at different values and see how our ACME changes for our Study 5 (pooled) sample. This requires the following assumed usual equations relating outcome (Y), treatment (T) and mediator (M) variables:

$$Y_i = \alpha_1 + \beta_1 T_i + \epsilon_{i1} \tag{1}$$

$$M_i = \alpha_2 + \beta_2 T_i + \epsilon_{i2} \tag{2}$$

$$Y_i = \alpha_3 + \beta_3 T_i + \gamma M_i + \epsilon_{i3} \tag{3}$$

We estimate that when ρ is around 0.12 the ACME becomes 0. Assume the unobserved

(pre-treatment) confounder formulation:

$$\epsilon_{i2} = \lambda_2 U_i + \epsilon'_{i2} \tag{4}$$

and

$$\epsilon_{i3} = \lambda_3 U_i + \epsilon'_{i3} \tag{5}$$

How much does U_i have to explain for our results to go away? Figure H.27 presents the proportion of original variance explained by U_i.

We can reparameterize ρ using $(\tilde{R}^2_M, \tilde{R}^2_Y)$:

$$\rho = \frac{\text{sgn}(\lambda_2\lambda_3)\tilde{R}_M\tilde{R}_Y}{\sqrt{(1-\tilde{R}^2_M)(1-\tilde{R}^2_Y)}} \tag{6}$$

where R^2_M and R^2_Y are from the original mediator/outcome models. We can set $(\tilde{R}^2_M, \tilde{R}^2_Y)$ to different values and see how mediation effects change.

Figure H.28 assumes that the confounder influences both the mediator and outcome variables in the same direction.[1] The bold line represents the various combinations of R^2 statistics where the ACME would be 0. In this case the product would have to be 0.014 for the ACME to become 0. Another way to say this is that when the product of the original variance explained by the omitted confounding is 0.014 the point estimate for ACME would be 0.

[1]This matters because the sensitivity analysis is in terms of the product of R^2 statistics; we assume positive because it seems more likely that something positively affecting the Mediator and the Outcome is happening to create the positive finding for the ACME).

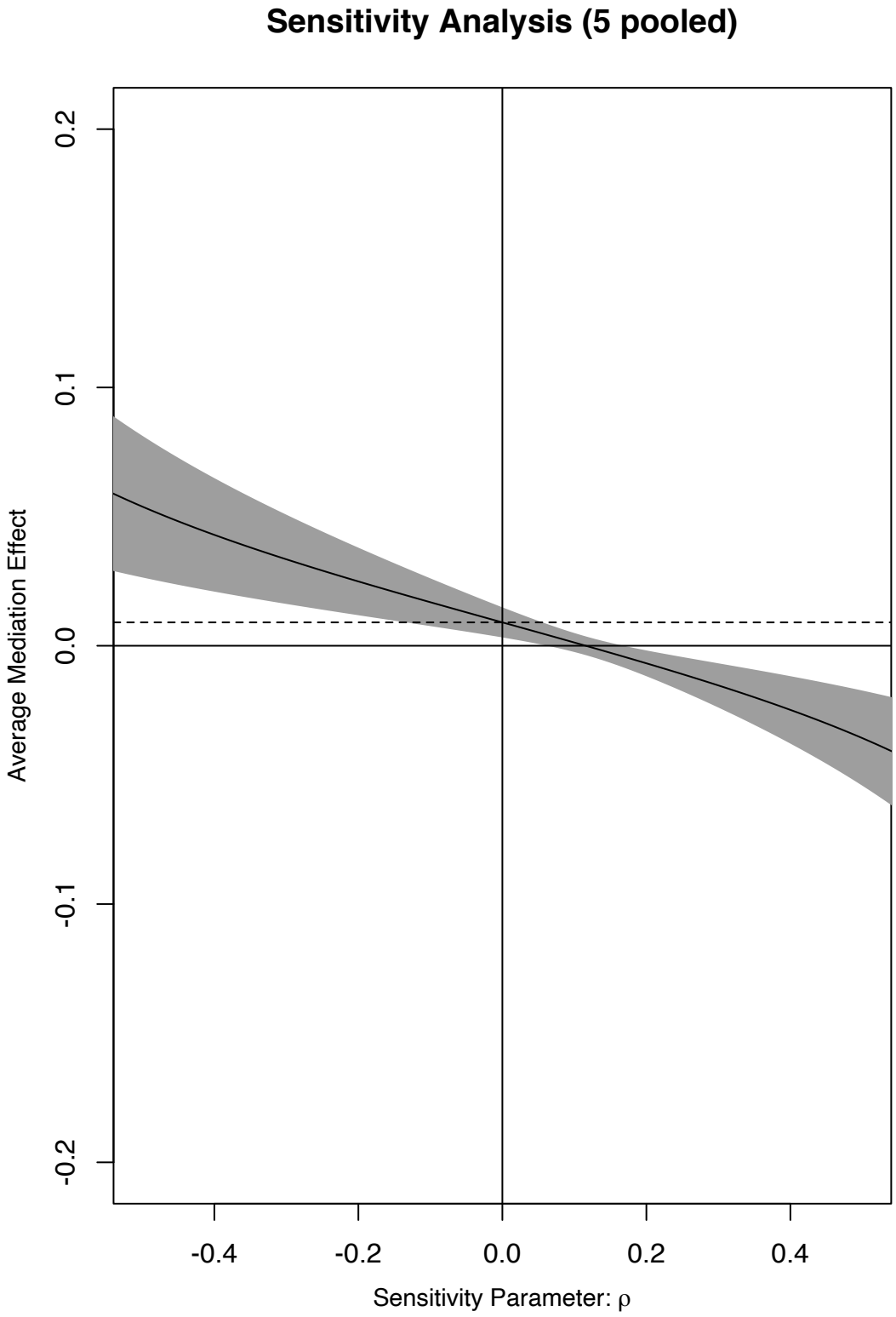


Figure H.27: **Proportion of original variance explained by U_i.**

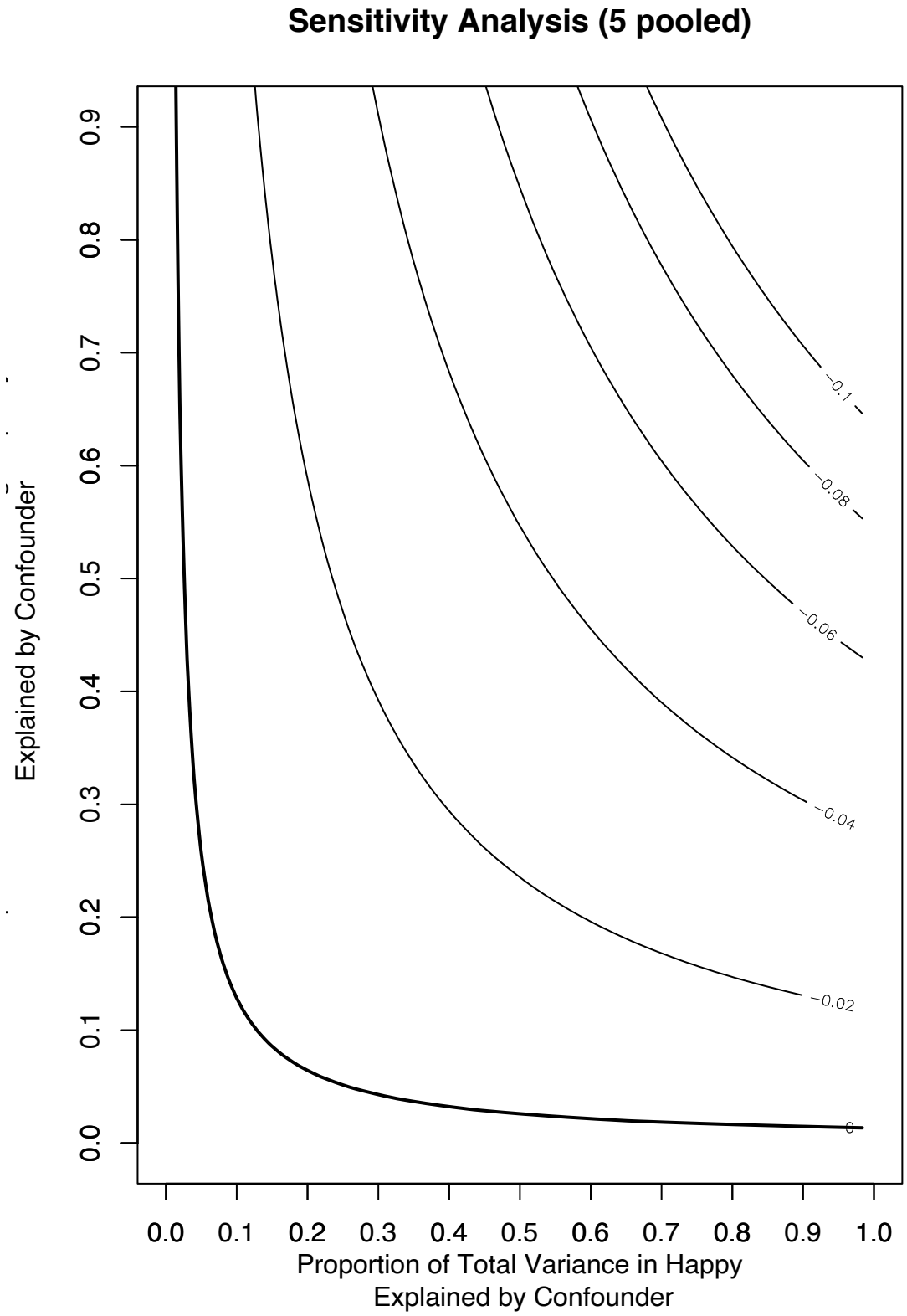


Figure H.28: R^2 **statistics for which ACME would be 0.**

I Scope of peer praise for empathy

Estimated standard errors are clustered at respondent levels and robust, and 90 and 95% confidence intervals are plotted throughout.

I.1 Subgroup analyses

by Party Democrats comprise of respondents who reported themselves as "Lean Democrat", "Democrat" and "Strong Democrat" while Republicans are respondents who reported themselves as "Lean Republican", "Republican" and "Strong Republican"; Independents are those who reported themselves as "Independent".

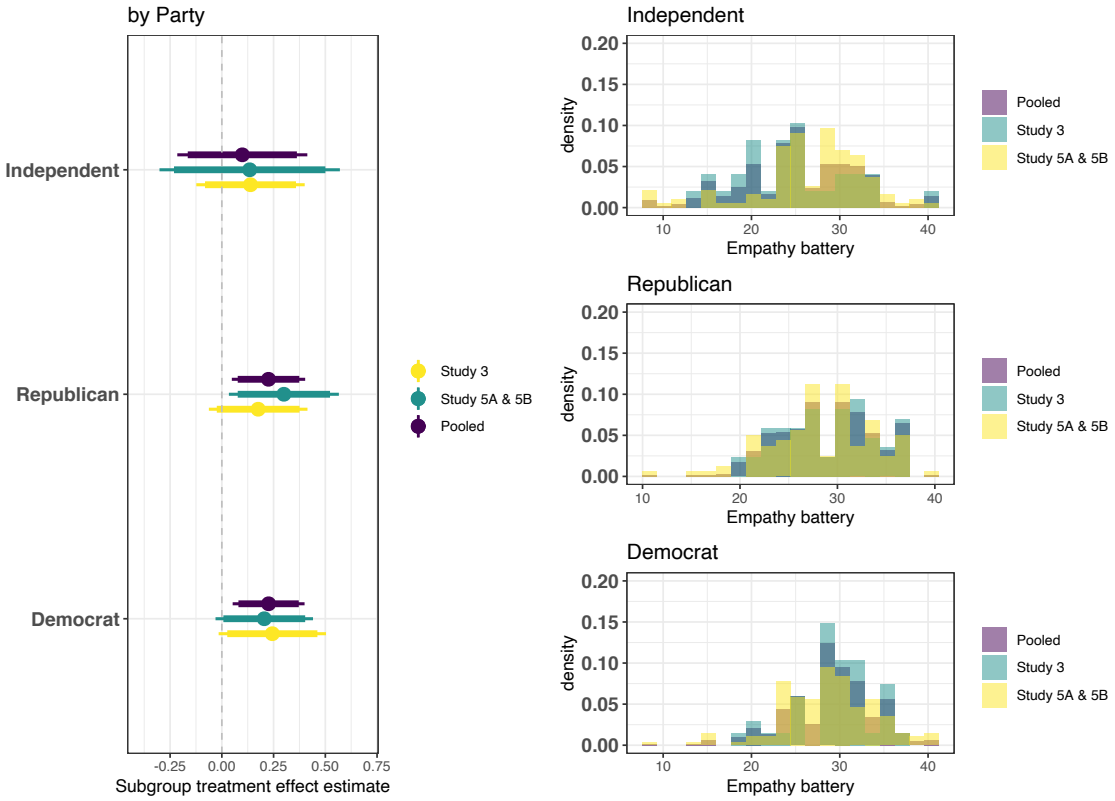


Figure I.29: Left panel: Logistic regression estimated peer praise for empathy treatment effect on log likelihood of choosing empathy task, by party subgroup. Right panel: density distribution of baseline empathy battery score by party subgroup.

by Trump and Biden approval Trump approval was measured in Study 3 under the question of presidential approval as Donald Trump was the then president-in-office; in Studies 5A and 5B to follow Joe Biden had taken office and so two questions were asked – one for presidential approval for Joe Biden, and a second on approval for former President Donald Trump. Figure I.30 presents subgroup analyses for Trump approval while Figure I.31 presents subgroup analyses for Biden approval.

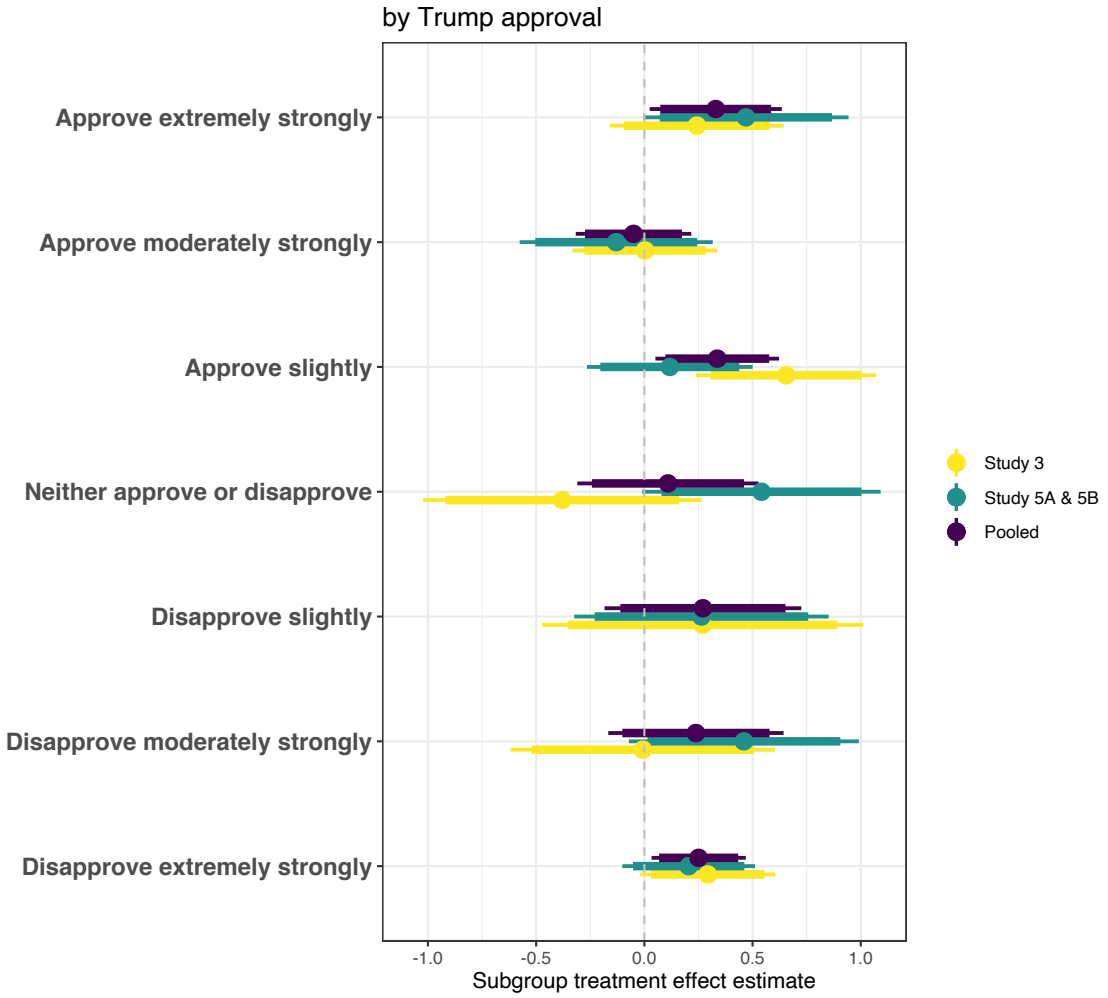


Figure I.30: Logistic regression estimated peer praise for empathy treatment effect on log likelihood of choosing empathy task, by Trump approval subgroup.

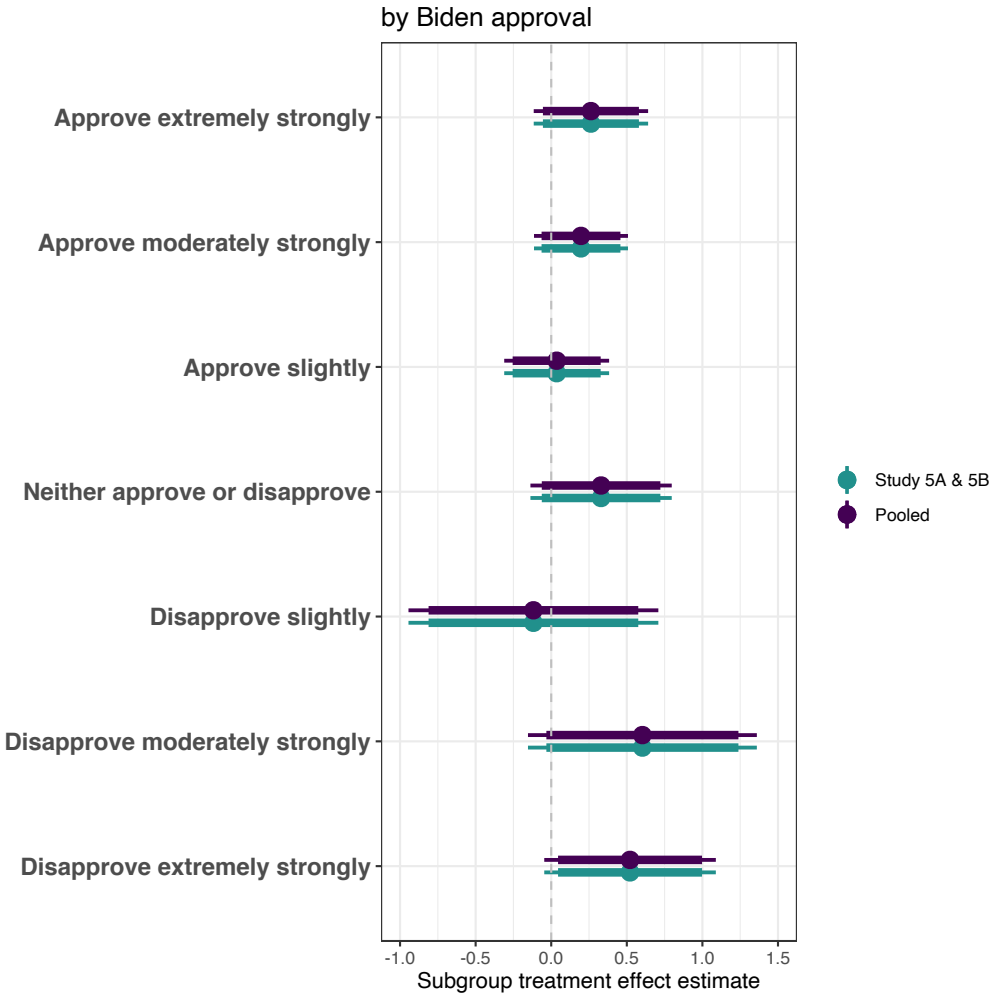


Figure I.31: Logistic regression estimated peer praise for empathy treatment effect on log likelihood of choosing empathy task, by Biden approval subgroup.

by Race Race subgroups presented are Black or African American ("Black"), White, and a combined grouping of Asian, Hispanic, Latinx, Native Hawaiian, Pacific Islander and Other ("Other") given the small sample sizes of the race subgroups. Figure I.32 presents subgroup analyses by race.

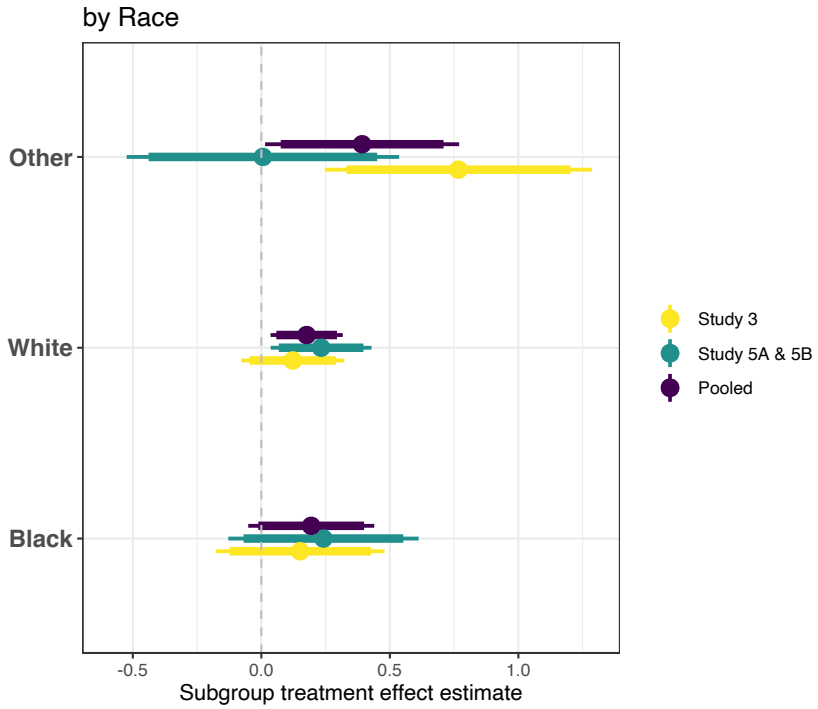


Figure I.32: Logistic regression estimated peer praise for empathy treatment effect on log likelihood of choosing empathy task, by race subgroup.

by Education Our most disaggregated coding for education level has too few observations for some categories for within-subgroup estimation of treatment effects, so we aggregate to three general categories, bundling ”Some high school, but did not graduate” and ”High school or equivalent (GED)” to ”HS”, combining ”Some college, but did not complete a degree” and ”Bachelor’s degree (BA/BS)” and ”Associate degree” to ”College”, and ”Master’s degree (MA/MS/MBA)” and ”Medical (MD), law (JD) or other doctoral degree (PhD)” combined to ”Postgrad”. In Study 5B we had an extra category for ”no schooling completed” but since this was a single respondent we drop this category throughout. Figure I.33 presents subgroup analyses by aggregated education level.

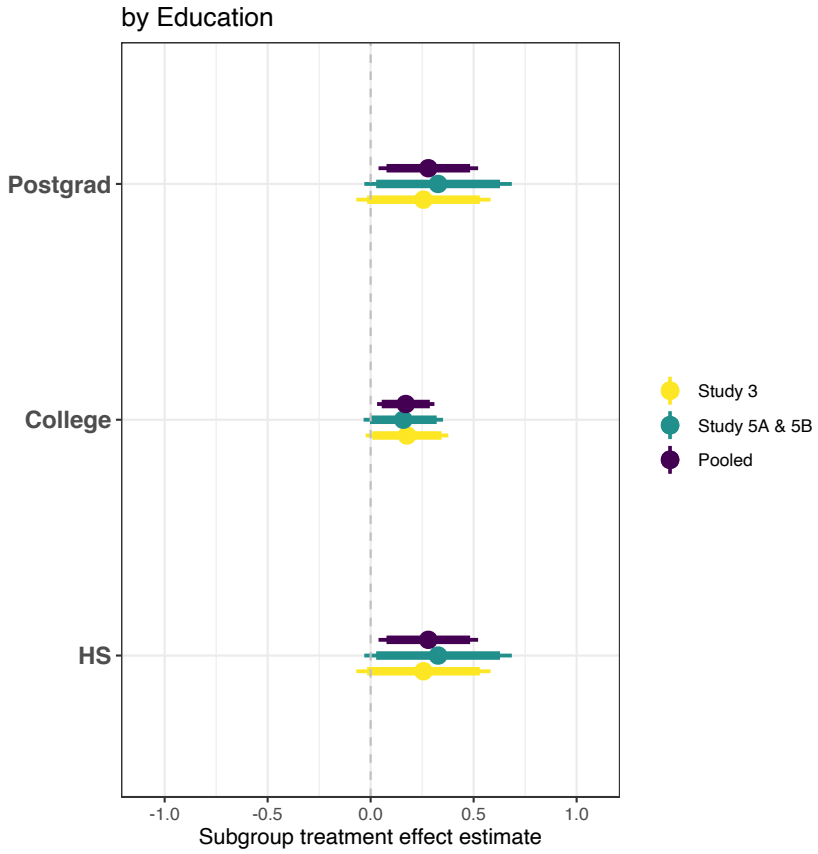


Figure I.33: Logistic regression estimated peer praise for empathy treatment effect on log likelihood of choosing empathy task, by education subgroup.

by Sex Figure I.34 presents subgroup analyses by respondent sex.

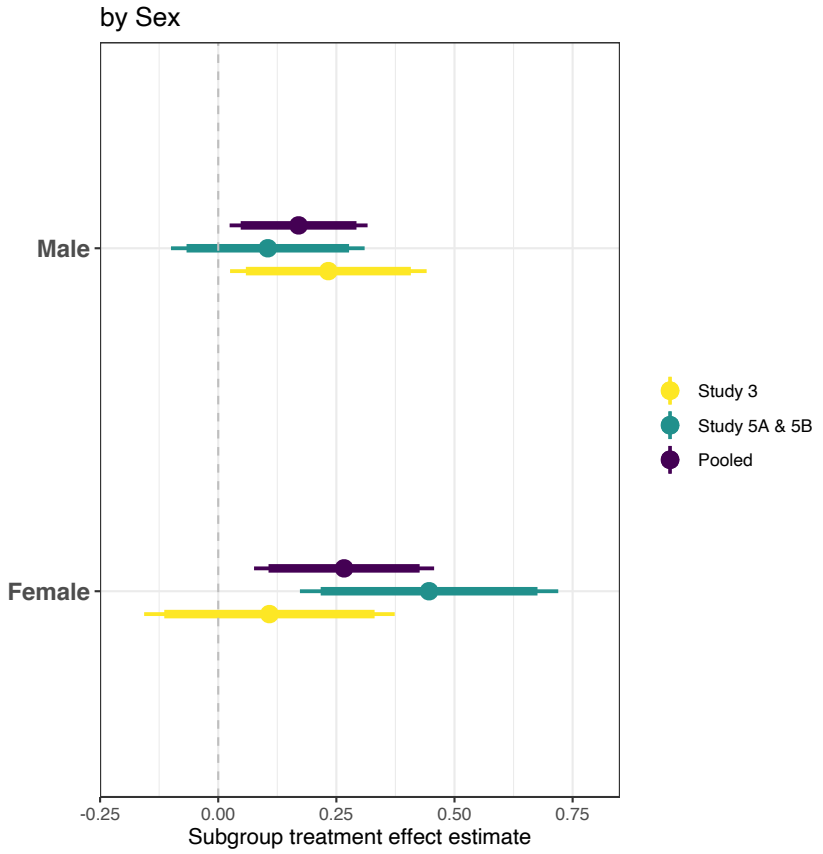


Figure I.34: Logistic regression estimated peer praise for empathy treatment effect on log likelihood of choosing empathy task, by respondent sex subgroup.

by baseline empathy In this section we consider subgroup effects of the peer praise treatment for baseline empathy. There are two reasons an exploration by baseline empathy is of interest; first, respondents who have high baseline measured *ability* or *practice* in being empathetic may be able to respond more to the treatment, but does this preclude individuals who score lower in this baseline? Second, measures of baseline empathy may also be capturing *proclivity* and general tendencies towards preferring empathy-oriented behaviors — a baseline "taste for empathy". Are we simply moving respondents who display such a taste alone using peer praise? In short, we find suggestive evidence against both arguments – while respondents who score in the highest tercile for baseline empathy are able to respond to the peer praise for empathy treatment, this does not preclude respondents from the middle tercile from doing so as well. Estimates of correlations between treatment and outcome by baseline empathy battery terciles (low, medium and high) presented in Figure I.35.

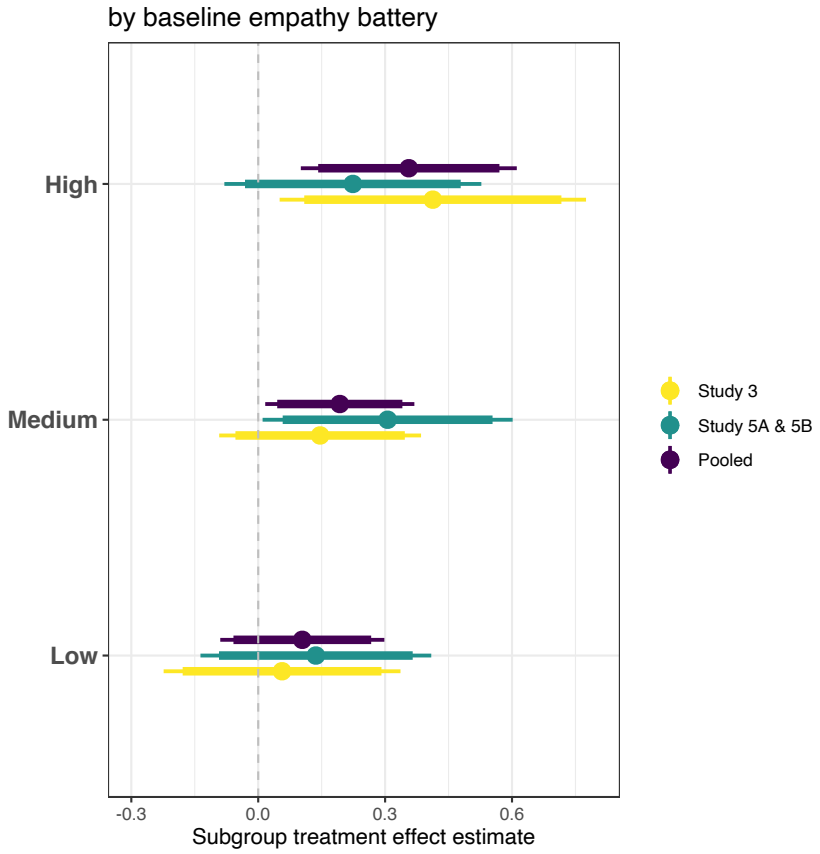


Figure I.35: Logistic regression estimated peer praise for empathy treatment effect on log likelihood of choosing empathy task, by respondent base empathy battery tercile.

by Attentiveness We look at subgroup effects by respondent attentiveness in Studies 5A and 5B (where the peer praise and task choice outcome are both measured for respondents over several trials) and look at attentive (respondents who pass the multiple choice `attentionMC` and grid `attentionG` attention checks), somewhat attentive (pass only `attentionMC` or `attentionG` but not both) and inattentive respondents (pass neither check). See Appendix Section B.4 for details. Figure I.36 presents estimated treatment effects of peer praise for empathy on choosing the empathy task within each of these subgroups.

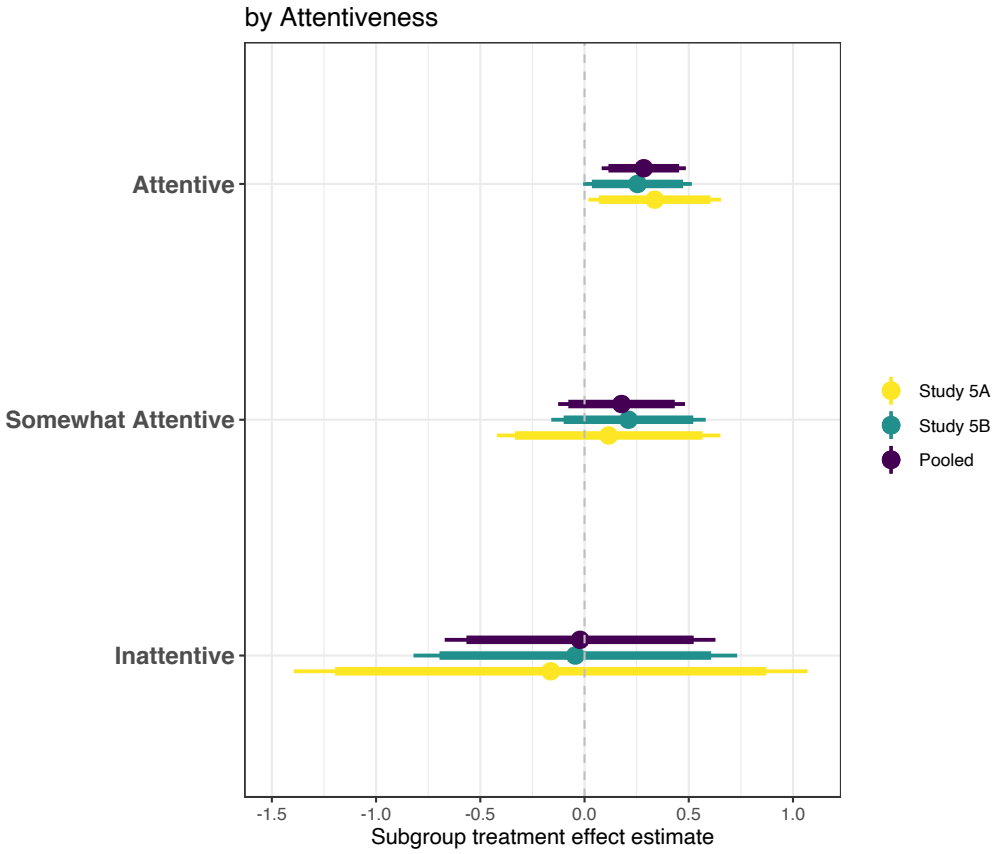


Figure I.36: Logistic regression estimated peer praise for empathy treatment effect on log likelihood of choosing empathy task, by respondent attentiveness.

Study 5 was composed of two days' worth of survey experiments, which we refer to throughout as 5A and 5B. 5A included 20 trials of the main task for all respondents, while 5B included 3. We additionally

I.2 Fading effects of peer praise

Figure I.37 presents estimated average causal marginal effects (ACME) and total effects (TE) or peer praise for empathy (through happiness) over successive main task trials.

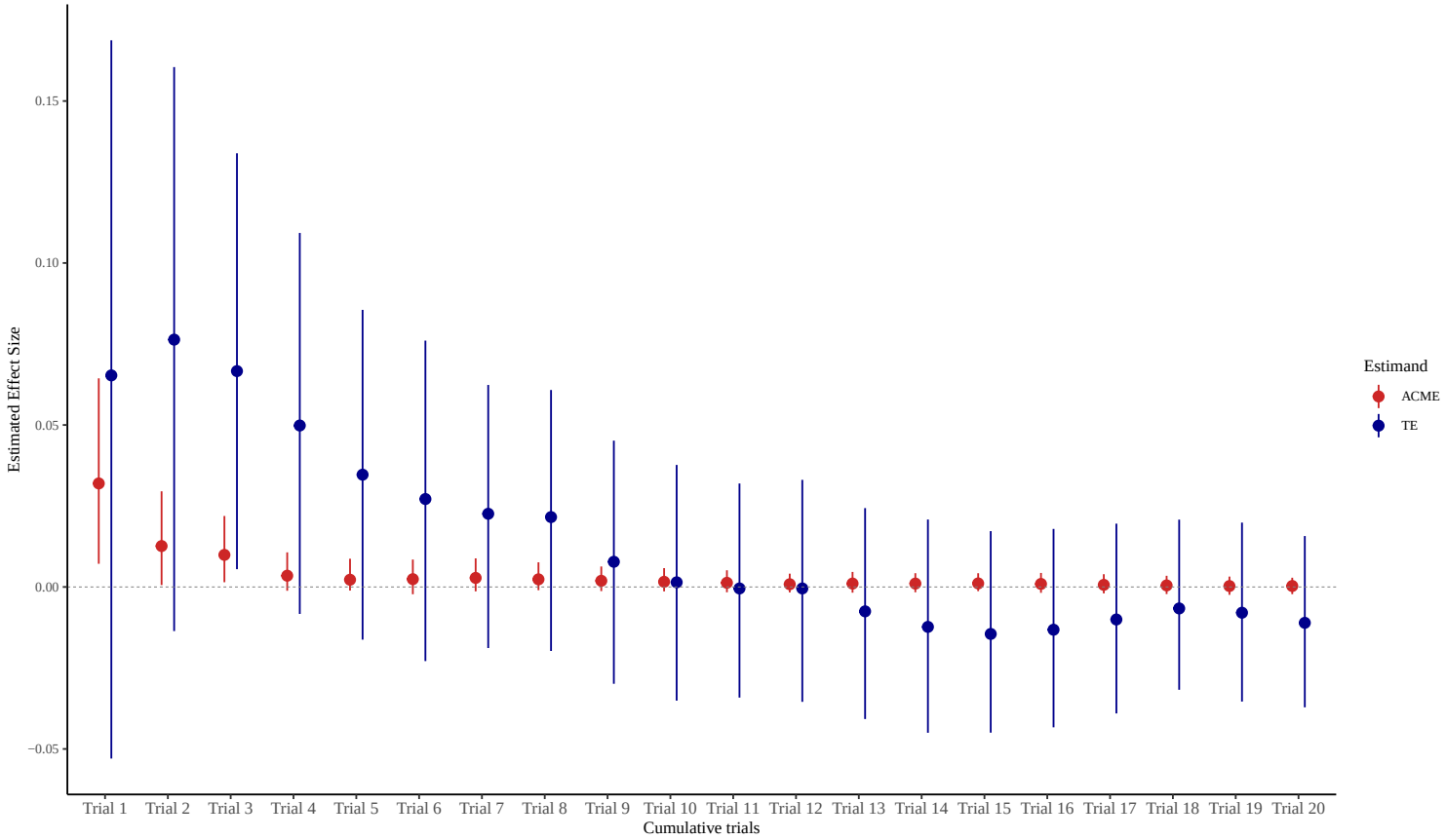


Figure I.37: **ACME and TE effects of peer praise (through happiness) over successive trials.**

J Explorations of other mechanisms for peer praise on empathy

In a short follow up study to Study 3 we also measured respondent reported anxiety randomly either before or after the tasks (one main task, one reservation wage task). We consider the difference in measured anxiety for control group respondents who chose the empathy task each time with the respondents who chose the objective task each time. If respondents become more anxious after choosing the empathy task then we should see that their change in anxiety values should be higher than their colleagues who chose the objective task each time (change in anxiety was measured as pre minus post). As this is an observational exploration, we further control for respondent sex, age, education, race and party. Table J.10 reports this analysis, and suggests that there is no evidence in support of the empathy task correlating with an increase in anxiety (coefficient on "Choose Empathy over Objective" is not significantly different from zero).

	Estimate	**s.e.**	**p**
Intercept	0.8892	0.1896	0.0000
Choose Empathy over Objective	-0.0224	0.0732	0.7593

Table J.10: Change in **anxiety** after choosing Empathy or Objective

To see if respondents alleviate anxiety of empathy through peer praise, we further explored changes in repoted anxiety among respondents who received either peer praise for empathy or control and who *only chose the empathy task* (comparing treatment effects on changes in anxiety while holding the task chosen constant). If it were the case that peer praise alleviates anxiety around the empathy task, we should see the peer praised group report higher changes in anxiety values compared to the second control. Again, this is an observational exploration, so we control for respondent sex, age, education, race and party. Table J.11 reports these findings, and suggests no evidence towards such a pathway.

	Estimate	**s.e.**	**p**
Intercept	0.5604	0.1623	0.0006
Peer praise	0.0589	0.0516	0.2536

Table J.11: Change in **anxiety** after peer praise for empathy

K Ethical considerations

All of the studies conducted in this project received IRB approval and exemption through the University of Wisconsin Madison Educational and Social/Behavioral Science IRB (# 2020-0843-CP002).

Fair wage In establishing pay scales for each study, we conducted pilots to establish average times for pre-treatment, task and post task portions of each study design and paid based on the state with the highest minimum wage in mid 2020 (Washington, at $13.50 per hour). Our intention was to offer fair wages especially in the context of work showing the median wage of MTurk workers is ˜$2/hour (Hara et al., 2018).

Negative treatments and distress In our studies we intentionally avoided negative affect in interactions as much as possible, by not providing negative peer feedback or emphasizing negative emotions when exploring mediators.

No deception Our studies incorporated a strict no-deception of respondents rule throughout, which in part motivated and necessitated Study 2 – garnering real peer praise and validating its authenticity.

References

Berinsky, Adam J, Michele F Margolis, Michael W Sances and Christopher Warshaw. 2019. "Using screeners to measure respondent attention on self-administered surveys: Which items and how many?" *Political Science Research and Methods* 9(2):1–8.

Hara, Kotaro, Abigail Adams, Kristy Milland, Saiph Savage, Chris Callison-Burch and Jeffrey P. Bigham. 2018. A Data-Driven Analysis of Workers' Earnings on Amazon Mechanical Turk. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. CHI '18 Montreal QC, Canada: Association for Computing Machinery pp. 1–14.
URL: *https://doi.org/10.1145/3173574.3174023*

Harmon-Jones, Cindy, Brock Bastian and Eddie Harmon-Jones. 2016. "The discrete emotions questionnaire: A new tool for measuring state self-reported emotions." *PloS one* 11(8):e0159915.

Imai, Kosuke, Luke Keele and Teppei Yamamoto. 2010. "Identification, Inference and Sensitivity Analysis for Causal Mediation Effects." *Statistical Science* 25(1):51–71.
URL: *https://www.jstor.org/stable/41058997*