Supplementary Material for Perspective-taking to Reduce Affective Polarization on Social Media

Martin Saveski[*]**, Nabeel Gillani**[*]**, Ann Yuan, Prashanth Vijayaraghavan, Deb Roy**
Massachusetts Institute of Technology

Contents

[*] Authors contributed equally

S1 Model Specification

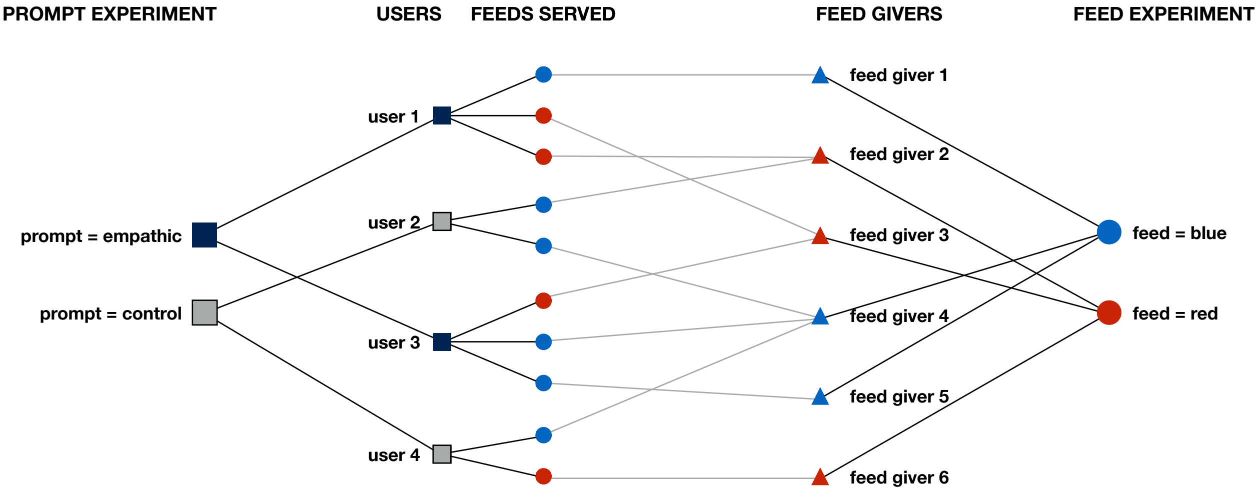

Figure S1: Illustration of the design explaining why the proposed mixed effects models are maximal. Note that (i) each user is assigned to only one prompt, (ii) one user may be served several red or blue feeds, (iii) each feed giver's feed is served to multiple users

As we discussed in the main text, we fit linear mixed-effects models with the maximal random effects structure justified by the design. Figure S1 shows a sketch of the experimental design and illustrates why the model specified in the main text is maximal. As the figure shows (i) each user is assigned to only one prompt, (ii) one user may be served several red or blue feeds, (iii) each feed giver's feed is served to multiple users. To account for these dependencies in the data, we include user random intercepts and random slopes, as well as feed giver random intercepts and random slopes:

$$y \sim feed * prompt + (1 + feed \mid user) + (1 + prompt \mid giver).$$

The user random intercepts and (feed) slopes allow the participants' responses to vary according to which feed condition they were assigned to, and feed giver random intercepts and (prompt) slopes allow the participants' responses to vary according to the prompt condition they are assigned to. (While here we focus on the random effects structure, the full model also includes interactions between the user covariates and the treatment assignments, as we discuss in the main text.)

We note that (a) the user random effects do not include prompt random slopes since the users are nested within the prompt condition (i.e., a user can either be in the control or empathic prompt condition and their response cannot vary by the prompt condition), and (b) the feed giver random effects do not include feed random slopes since the destination feed owners are nested within the feed condition (i.e., a destination feed is either the same or opposing political ideology as the participant, and only one destination feed owner + feed combination is possible).

S2 Robustness Analyses

To test the robustness of the results reported in the main text, we rerun the analyses using different model specifications. We test eight different model specifications, varying three aspects of the models: (*i*) whether the model includes a term interacting the two treatment assignments (feed and prompt), (*ii*) whether the model includes the user covariates, and (*iii*) whether the model is frequentist (*lmer* [1]) or Bayesian (*brms* [2, 3]). Including an interaction term allows us to model how the two treatments interact with each other. For instance, the empathic prompt might be more effective when the user is shown a feed of someone with an opposite political leaning. Including the user covariates allows us to adjust for any imbalance between the treatment and control groups and potentially gain statistical power. In all model specifications that include the covariates, we interact each of the covariates with the two treatment assignments. Finally, fitting Bayesian models (with *brms*) allowed us to fit the maximal models, including the random slopes. In contrast, fitting frequentist models (with *lmer*) with the maximal specification in almost all cases did not converge, and we had to simplify the model until we found one that converges, as recommended in the literature [4]. All *lmer* models reported below include only user and giver random intercepts. To fit the Bayesian models, we used $\mathcal{N}(0, 30)$ priors for the coefficients and the intercepts, and $\mathcal{N}(0, 1)$ for the standard deviations. For each model, we run four chains with 10,000 iterations each. Code listing 1 shows the R code used to fit the different models using Time Spent as an example outcome.

Figure S2 shows the estimates and credible/confidence intervals (CIs) for the intercept and the feed and prompt treatment assignments. The complete model outputs are shown in Tables S1-S12. In most cases, all eight model specifications lead to the same substantive conclusions. The estimates and CIs of the corresponding *brms* and *lmer* models are very close. Similarly, including the covariates in the models leads to very small changes in the estimates. The most consequential modeling decision is whether to include a term that interacts the two treatment assignments. While including an interaction term does not drastically affect the point estimates, it does lead to significantly wider CIs—this is particularly the case for the CIs of the prompt term for the survey question responses. We note that the model specification reported in the main text (*brms*, with interactions, with covariates) leads to the most conservative results.

Listing 1: R code used to fit the different model specifications, using *Time Spent* as an example outcome.

```r
#
# brms settings
#
n_chains <- 4
n_iters <- 10000
n_cores <- 4
c_adapt_delta <- 0.99
c_max_treedepth <- 15
priors <- c(
  prior_string("normal(0, 30)", class = "Intercept"),
  prior_string("normal(0, 30)", class = "b"),
  prior_string("normal(0, 1)", class = "sd")
)

#
# Model specifications
#

# lmer
ts_lmer   <- lmer(timespent ~ feed_opp + prompt + (1 | user) + (1 | giver), data=df)
```

```
# lmer, with interactions
ts_i_lmer <- lmer(timespent ~ feed_opp * prompt + (1 | user) + (1 | giver), data=df)

# brms
ts_brms <- brm(
  timespent ~ feed_opp + prompt + (1 + feed_opp | user) + (1 + prompt | giver),
  data = df,
  family = gaussian(),
  chains = n_chains, iter = n_iters, cores = n_cores, seed = 0,
  control = list(adapt_delta = c_adapt_delta, max_treedepth = c_max_treedepth),
  prior = priors
)

# brms, with interactions
ts_i_brms <- brm(
  timespent ~ feed_opp * prompt + (1 + feed_opp | user) + (1 + prompt | giver),
  data = df,
  family = gaussian(),
  chains = n_chains, iter = n_iters, cores = n_cores, seed = 0,
  control = list(adapt_delta = c_adapt_delta, max_treedepth = c_max_treedepth),
  prior = priors
)

# lmer, with covariates
ts_lmer <- lmer(
  timespent ~
    feed_opp + prompt +
    feed_opp * days_active +
    feed_opp * num_statuses +
    feed_opp * num_favorites +
    feed_opp * num_followers +
    feed_opp * num_friends +
    prompt * days_active +
    prompt * num_statuses +
    prompt * num_favorites +
    prompt * num_followers +
    prompt * num_friends +
    (1 | user) +
    (1 | giver),
  data = df
)

# lmer, with interactions, with covariates
ts_i_lmer <- lmer(
  timespent ~
    feed_opp * prompt +
    feed_opp * days_active +
    feed_opp * num_statuses +
    feed_opp * num_favorites +
    feed_opp * num_followers +
    feed_opp * num_friends +
    prompt * days_active +
    prompt * num_statuses +
    prompt * num_favorites +
    prompt * num_followers +
    prompt * num_friends +
    (1 | user) +
    (1 | giver),
  data = df
)

# brms, with covariates
ts_brms <- brm(
  timespent ~
    feed_opp + prompt +
    feed_opp * days_active +
    feed_opp * num_statuses +
    feed_opp * num_favorites +
    feed_opp * num_followers +
    feed_opp * num_friends +
    prompt * days_active +
    prompt * num_statuses +
    prompt * num_favorites +
    prompt * num_followers +
    prompt * num_friends +
    (1 + feed_opp | user) +
    (1 + prompt | giver),
  data = df,
  family = gaussian(),
  chains = n_chains, iter = n_iters, cores = n_cores, seed = 0,
  control = list(adapt_delta = c_adapt_delta, max_treedepth = c_max_treedepth),
  prior = priors
)

# brms, with interactions, with covariates
ts_i_brms <- brm(
  timespent ~
    feed_opp * prompt +
    feed_opp * days_active +
    feed_opp * num_statuses +
    feed_opp * num_favorites +
    feed_opp * num_followers +
    feed_opp * num_friends +
    prompt * days_active +
    prompt * num_statuses +
    prompt * num_favorites +
    prompt * num_followers +
    prompt * num_friends +
    (1 + feed_opp | user) +
    (1 + prompt | giver),
  data = df,
  family = gaussian(),
  chains = n_chains, iter = n_iters, cores = n_cores, seed = 0,
  control = list(adapt_delta = c_adapt_delta, max_treedepth = c_max_treedepth),
  prior = priors
)
```

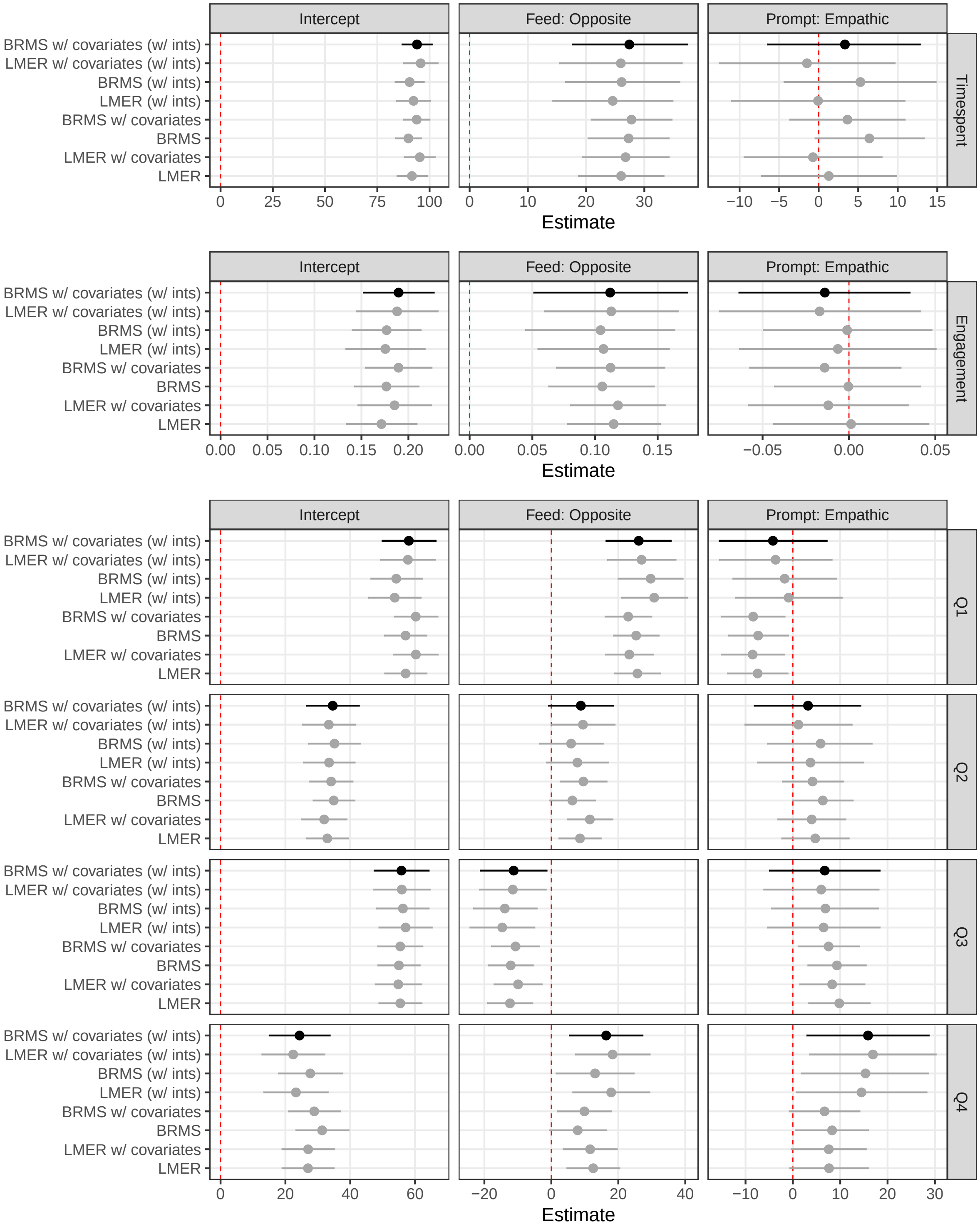

Figure S2: Analysis using eight different model specifications. The points represent point estimates, and the error bars represent 95% credible/confidence intervals. The figure shows only the key estimates, the complete model outputs are shown in Tables S1-S12. The highlighted model (black) is the one reported in the main text.

	Outcome: Time Spent (Seconds)			
	BRMS	*BRMS* *w/ interactions*	*LMER*	*LMER* *w/ interactions*
(Intercept)	89.87*	90.46*	91.57*	92.31*
	[83.63; 96.15]	[83.36; 97.50]	[84.22; 98.93]	[84.07; 100.56]
feed=opp	27.30*	26.11*	26.01*	24.57*
	[20.27; 34.30]	[16.37; 36.11]	[18.64; 33.39]	[14.21; 34.93]
prompt=emp	6.41	5.28	1.28	−0.08
	[−0.50; 13.35]	[−4.42; 14.90]	[−7.31; 9.88]	[−11.08; 10.92]
feed=opp × *prompt=emp*		2.37		2.73
		[−11.26; 15.80]		[−11.02; 16.47]
Num. obs.	6687.00	6687.00	6687.00	6687.00

* 0 outside the credible/confidence interval

Table S1: Results of the analysis of *Time Spent* using different model specifications: with vs. without a treatments interaction term, and using BRMS vs. LMER.

	Outcome: Time Spent (Seconds)			
	BRMS	*BRMS* *w/ interactions*	*LMER*	*LMER* *w/ interactions*
(Intercept)	93.84*	93.99*	95.33*	95.75*
	[87.43; 100.19]	[86.64; 101.31]	[87.76; 102.90]	[87.27; 104.22]
feed=opp	27.79*	27.42*	26.78*	25.96*
	[20.82; 34.78]	[17.56; 37.40]	[19.25; 34.30]	[15.41; 36.52]
prompt=emp	3.64	3.32	−0.72	−1.48
	[−3.70; 10.95]	[−6.48; 12.91]	[−9.49; 8.05]	[−12.65; 9.69]
feed=opp × *prompt=emp*		0.79		1.52
		[−12.78; 14.01]		[−12.30; 15.33]
days active	−1.13	−1.17	−1.58	−1.61
	[−8.41; 6.07]	[−8.45; 6.14]	[−9.89; 6.74]	[−9.93; 6.71]
statuses count	−6.19	−6.10	−7.14	−7.14
	[−17.53; 5.24]	[−17.52; 5.26]	[−20.39; 6.12]	[−20.40; 6.12]
favorites count	−10.87*	−10.90*	−9.24	−9.28
	[−19.32; −2.37]	[−19.48; −2.44]	[−19.02; 0.53]	[−19.06; 0.49]
followers count	−4.35	−4.31	−5.57	−5.53
	[−14.88; 6.10]	[−14.78; 6.29]	[−17.91; 6.78]	[−17.88; 6.81]
friends count	7.97	7.89	8.73	8.70
	[−0.74; 16.77]	[−0.75; 16.68]	[−1.46; 18.91]	[−1.49; 18.89]
feed=opp × *days active*	2.07	2.12	1.37	1.41
	[−6.10; 10.26]	[−6.26; 10.38]	[−6.90; 9.63]	[−6.87; 9.68]
feed=opp × *statuses count*	2.13	2.13	2.76	2.76
	[−10.73; 14.87]	[−10.75; 14.99]	[−10.48; 16.01]	[−10.49; 16.00]
feed=opp × *favorites count*	−5.11	−5.07	−7.70	−7.61
	[−14.72; 4.55]	[−14.83; 4.65]	[−17.47; 2.07]	[−17.42; 2.19]
feed=opp × *followers count*	−2.24	−2.32	−2.05	−2.08
	[−14.00; 9.62]	[−14.24; 9.46]	[−14.24; 10.14]	[−14.28; 10.11]
feed=opp × *friends count*	0.85	0.90	1.71	1.73
	[−9.16; 10.84]	[−9.30; 11.09]	[−8.50; 11.92]	[−8.49; 11.94]
prompt=emp × *days active*	−0.75	−0.75	0.84	0.86
	[−9.09; 7.54]	[−9.06; 7.65]	[−9.31; 11.00]	[−9.30; 11.02]
prompt=emp × *statuses count*	10.43	10.25	3.94	3.94
	[−2.61; 23.42]	[−2.56; 23.23]	[−12.48; 20.36]	[−12.48; 20.37]
prompt=emp × *favorites count*	−10.05*	−10.01*	2.23	2.24
	[−19.84; −0.36]	[−19.77; −0.30]	[−10.19; 14.65]	[−10.18; 14.66]
prompt=emp × *followers count*	−2.30	−2.23	−0.87	−0.89
	[−14.26; 9.46]	[−14.20; 9.70]	[−15.72; 13.98]	[−15.75; 13.96]
prompt=emp × *friends count*	−3.20	−3.17	−5.34	−5.34
	[−13.39; 7.00]	[−13.10; 6.85]	[−17.73; 7.05]	[−17.73; 7.05]
Num. obs.	6687.00	6687.00	6687.00	6687.00

* 0 outside the credible/confidence interval

Table S2: Results of the analysis of *Time Spent* using different model specifications: with vs. without a treatments interaction term, and using BRMS vs. LMER.

	Outcome: Engagement			
	BRMS	*BRMS* *w/ interactions*	*LMER*	*LMER* *w/ interactions*
(Intercept)	0.18^{*}	0.18^{*}	0.17^{*}	0.18^{*}
	[0.14; 0.21]	[0.14; 0.21]	[0.13; 0.21]	[0.13; 0.22]
feed=opp	0.11^{*}	0.10^{*}	0.11^{*}	0.11^{*}
	[0.06; 0.15]	[0.04; 0.16]	[0.08; 0.15]	[0.05; 0.16]
prompt=emp	−0.00	−0.00	0.00	−0.01
	[−0.04; 0.04]	[−0.05; 0.05]	[−0.04; 0.05]	[−0.06; 0.05]
feed=opp × *prompt=emp*		0.00		0.02
		[−0.08; 0.09]		[−0.05; 0.09]
Num. obs.	6687.00	6687.00	6687.00	6687.00

* 0 outside the credible/confidence interval

Table S3: Results of the analysis of *Engagement* using different model specifications: with vs. without a treatments interaction term, and using BRMS vs. LMER.

	Outcome: Engagement			
	BRMS	*BRMS w/ interactions*	*LMER*	*LMER w/ interactions*
(Intercept)	0.19*	0.19*	0.19*	0.19*
	[0.15; 0.23]	[0.15; 0.23]	[0.15; 0.22]	[0.14; 0.23]
feed=opp	0.11*	0.11*	0.12*	0.11*
	[0.07; 0.16]	[0.05; 0.17]	[0.08; 0.16]	[0.06; 0.17]
prompt=emp	−0.01	−0.01	−0.01	−0.02
	[−0.06; 0.03]	[−0.06; 0.04]	[−0.06; 0.03]	[−0.08; 0.04]
feed=opp × *prompt=emp*		0.00		0.01
		[−0.08; 0.08]		[−0.06; 0.08]
days active	0.02	0.02	0.01	0.01
	[−0.02; 0.05]	[−0.02; 0.05]	[−0.03; 0.06]	[−0.03; 0.06]
statuses count	−0.06	−0.06*	−0.07*	−0.07*
	[−0.12; 0.00]	[−0.12; −0.00]	[−0.14; −0.00]	[−0.14; −0.00]
favorites count	0.00	0.00	0.01	0.01
	[−0.04; 0.05]	[−0.04; 0.05]	[−0.04; 0.06]	[−0.04; 0.06]
followers count	−0.01	−0.01	−0.01	−0.01
	[−0.06; 0.05]	[−0.06; 0.05]	[−0.08; 0.05]	[−0.08; 0.05]
friends count	0.00	0.00	0.01	0.01
	[−0.05; 0.05]	[−0.04; 0.05]	[−0.04; 0.07]	[−0.04; 0.07]
feed=opp × *days active*	−0.01	−0.01	−0.02	−0.02
	[−0.06; 0.04]	[−0.06; 0.04]	[−0.06; 0.02]	[−0.06; 0.02]
feed=opp × *statuses count*	−0.02	−0.02	0.00	0.00
	[−0.10; 0.05]	[−0.10; 0.06]	[−0.07; 0.07]	[−0.07; 0.07]
feed=opp × *favorites count*	−0.01	−0.01	−0.03	−0.02
	[−0.07; 0.05]	[−0.07; 0.05]	[−0.08; 0.02]	[−0.07; 0.03]
feed=opp × *followers count*	0.02	0.02	0.02	0.02
	[−0.05; 0.09]	[−0.05; 0.09]	[−0.04; 0.08]	[−0.04; 0.08]
feed=opp × *friends count*	−0.02	−0.02	−0.02	−0.02
	[−0.08; 0.04]	[−0.08; 0.04]	[−0.07; 0.04]	[−0.07; 0.04]
prompt=emp × *days active*	−0.03	−0.03	−0.02	−0.02
	[−0.08; 0.02]	[−0.08; 0.02]	[−0.07; 0.03]	[−0.07; 0.03]
prompt=emp × *statuses count*	0.04	0.04	0.04	0.04
	[−0.04; 0.12]	[−0.04; 0.12]	[−0.04; 0.13]	[−0.04; 0.13]
prompt=emp × *favorites count*	0.03	0.03	0.02	0.02
	[−0.03; 0.09]	[−0.03; 0.09]	[−0.05; 0.09]	[−0.05; 0.09]
prompt=emp × *followers count*	0.01	0.01	0.02	0.02
	[−0.06; 0.09]	[−0.06; 0.08]	[−0.05; 0.10]	[−0.05; 0.10]
prompt=emp × *friends count*	−0.02	−0.02	−0.03	−0.03
	[−0.08; 0.05]	[−0.08; 0.04]	[−0.10; 0.03]	[−0.10; 0.03]
Num. obs.	6687.00	6687.00	6687.00	6687.00

* 0 outside the credible/confidence interval

Table S4: Results of the analysis of *Engagement* using different model specifications: with vs. without a treatments interaction term, and using BRMS vs. LMER.

	Outcome: Survey Q1 (”This feed is different from what I’m used to seeing”)			
	BRMS	*BRMS* *w/ interactions*	*LMER*	*LMER* *w/ interactions*
(Intercept)	57.08*	54.19*	57.11*	53.72*
	[50.47; 63.66]	[46.25; 62.27]	[50.53; 63.68]	[45.56; 61.88]
feed=opp	25.33*	29.67*	25.70*	30.73*
	[18.49; 32.26]	[19.86; 39.37]	[18.83; 32.58]	[20.80; 40.66]
prompt=emp	−7.34*	−1.74	−7.45*	−0.91
	[−13.67; −0.90]	[−12.75; 9.28]	[−13.87; −1.02]	[−12.24; 10.42]
feed=opp × *prompt=emp*		−8.34		−9.62
		[−21.53; 4.82]		[−23.37; 4.12]
Num. obs.	253.00	253.00	253.00	253.00

* 0 outside the credible/confidence interval

Table S5: Results of the analysis of *Survey Question 1* using different model specifications: with vs. without a treatments interaction term, and using BRMS vs. LMER.

	Outcome: Survey Q1			
	(”This feed is different from what I’m used to seeing”)			
	BRMS	*BRMS* *w/ interactions*	*LMER*	*LMER* *w/ interactions*
(Intercept)	60.22*	58.08*	60.26*	57.79*
	[53.42; 67.07]	[49.69; 66.53]	[53.32; 67.20]	[49.24; 66.35]
feed=opp	22.96*	26.12*	23.28*	26.96*
	[15.92; 30.01]	[16.23; 35.91]	[16.11; 30.46]	[16.68; 37.24]
prompt=emp	−8.39*	−4.21	−8.50*	−3.65
	[−15.13; −1.65]	[−15.67; 7.32]	[−15.22; −1.78]	[−15.59; 8.29]
feed=opp × *prompt=emp*		−6.04		−6.98
		[−19.47; 7.48]		[−21.10; 7.14]
days active	−4.94	−5.22	−5.33	−5.58
	[−12.74; 2.98]	[−13.00; 2.56]	[−13.40; 2.75]	[−13.68; 2.53]
statuses count	−5.70	−5.46	−5.51	−5.04
	[−19.56; 8.34]	[−19.53; 8.73]	[−20.44; 9.43]	[−19.98; 9.90]
favorites count	−5.68	−5.88	−6.08	−6.42
	[−15.88; 4.60]	[−16.05; 4.32]	[−16.79; 4.63]	[−17.15; 4.32]
followers count	2.54	3.27	2.97	3.74
	[−7.30; 12.61]	[−6.76; 13.28]	[−7.22; 13.15]	[−6.58; 14.07]
friends count	−7.83	−7.74	−8.03	−8.08
	[−17.60; 1.67]	[−17.47; 2.10]	[−18.10; 2.04]	[−18.16; 2.00]
feed=opp × *days active*	2.57	2.68	2.89	2.97
	[−5.26; 10.32]	[−5.15; 10.40]	[−5.09; 10.87]	[−5.02; 10.96]
feed=opp × *statuses count*	−1.13	−1.58	−1.55	−2.28
	[−15.58; 13.55]	[−16.26; 12.78]	[−17.07; 13.97]	[−17.86; 13.30]
feed=opp × *favorites count*	12.00*	12.14*	12.40*	12.73*
	[1.02; 23.02]	[1.15; 23.24]	[0.90; 23.89]	[1.21; 24.25]
feed=opp × *followers count*	1.16	0.47	0.90	0.04
	[−10.25; 12.66]	[−11.18; 11.82]	[−10.87; 12.68]	[−11.87; 11.95]
feed=opp × *friends count*	5.97	6.12	6.27	6.64
	[−4.86; 16.93]	[−4.92; 17.23]	[−5.08; 17.62]	[−4.73; 18.00]
prompt=emp × *days active*	5.72	5.89	5.93	6.08
	[−1.55; 12.94]	[−1.32; 13.19]	[−1.46; 13.31]	[−1.31; 13.47]
prompt=emp × *statuses count*	1.68	1.81	1.86	1.94
	[−11.67; 14.90]	[−11.59; 14.88]	[−11.83; 15.56]	[−11.76; 15.65]
prompt=emp × *favorites count*	2.29	2.11	2.45	2.21
	[−8.01; 12.53]	[−8.22; 12.36]	[−8.17; 13.07]	[−8.42; 12.84]
prompt=emp × *followers count*	0.10	0.18	−0.25	−0.08
	[−11.33; 11.57]	[−11.44; 11.56]	[−11.91; 11.41]	[−11.75; 11.58]
prompt=emp × *friends count*	0.96	0.70	0.91	0.67
	[−9.56; 11.59]	[−9.96; 11.41]	[−9.81; 11.63]	[−10.08; 11.41]
Num. obs.	253.00	253.00	253.00	253.00

* 0 outside the credible/confidence interval

Table S6: Results of the analysis of *Survey Question 1* using different model specifications: with vs. without a treatments interaction term, and using BRMS vs. LMER.

	Outcome: Survey Q2			
	(”I learned something new from browsing this person’s feed”)			
	BRMS	*BRMS w/ interactions*	*LMER*	*LMER w/ interactions*
(Intercept)	34.91*	35.14*	32.93*	33.47*
	[28.44; 41.46]	[27.04; 43.24]	[26.30; 39.56]	[25.42; 41.52]
feed=opp	6.29	5.91	8.56*	7.78
	[−0.56; 13.19]	[−3.66; 15.60]	[2.16; 14.97]	[−1.60; 17.16]
prompt=emp	6.33	5.86	4.74	3.72
	[−0.14; 12.77]	[−5.45; 16.81]	[−2.43; 11.90]	[−7.49; 14.94]
feed=opp × *prompt=emp*		0.69		1.44
		[−12.59; 14.25]		[−10.90; 13.79]
Num. obs.	246.00	246.00	246.00	246.00

* 0 outside the credible/confidence interval

Table S7: Results of the analysis of *Survey Question 2* using different model specifications: with vs. without a treatments interaction term, and using BRMS vs. LMER.

	Outcome: Survey Q2 (”I learned something new from browsing this person’s feed”)			
	BRMS	*BRMS* *w/ interactions*	*LMER*	*LMER* *w/ interactions*
(Intercept)	34.10*	34.60*	31.96*	33.39*
	[27.41; 40.86]	[26.39; 42.85]	[24.94; 38.98]	[25.05; 41.74]
feed=opp	9.57*	8.82	11.51*	9.44
	[2.50; 16.65]	[−0.92; 18.54]	[4.63; 18.39]	[−0.18; 19.07]
prompt=emp	4.16	3.20	3.98	1.18
	[−2.31; 10.76]	[−8.24; 14.37]	[−3.26; 11.21]	[−10.23; 12.58]
feed=opp × *prompt=emp*		1.42		3.97
		[−11.73; 14.85]		[−8.58; 16.51]
days active	−7.17	−7.01	−5.67	−5.57
	[−14.89; 0.60]	[−14.83; 0.91]	[−13.48; 2.15]	[−13.41; 2.27]
statuses count	10.20	10.07	5.56	5.54
	[−3.67; 24.07]	[−3.85; 23.95]	[−9.08; 20.19]	[−9.13; 20.22]
favorites count	−3.64	−3.46	0.39	0.49
	[−13.77; 6.49]	[−13.63; 6.53]	[−9.99; 10.77]	[−9.93; 10.91]
followers count	1.63	1.41	0.03	−0.14
	[−8.14; 11.39]	[−8.42; 11.28]	[−10.73; 10.79]	[−10.94; 10.65]
friends count	−10.98*	−11.05*	−8.90	−9.12
	[−20.84; −1.08]	[−20.79; −1.04]	[−19.09; 1.29]	[−19.35; 1.11]
feed=opp × *days active*	5.23	5.13	3.46	3.55
	[−2.56; 13.09]	[−2.61; 12.82]	[−3.44; 10.37]	[−3.37; 10.46]
feed=opp × *statuses count*	−18.00*	−17.79*	−10.96	−10.79
	[−32.65; −3.46]	[−32.52; −2.97]	[−25.46; 3.55]	[−25.36; 3.78]
feed=opp × *favorites count*	−0.65	−0.84	−2.49	−2.74
	[−11.36; 10.29]	[−11.84; 10.09]	[−12.80; 7.81]	[−13.11; 7.63]
feed=opp × *followers count*	3.20	3.35	2.51	2.61
	[−8.18; 14.59]	[−8.00; 14.66]	[−8.64; 13.65]	[−8.56; 13.78]
feed=opp × *friends count*	10.87	10.93	6.62	6.70
	[−0.21; 21.97]	[−0.31; 21.95]	[−3.79; 17.03]	[−3.74; 17.14]
prompt=emp × *days active*	1.30	1.20	1.50	1.39
	[−5.91; 8.58]	[−6.02; 8.29]	[−6.68; 9.69]	[−6.83; 9.60]
prompt=emp × *statuses count*	−11.11	−11.15	−11.56	−11.83
	[−24.31; 2.02]	[−24.40; 2.33]	[−26.58; 3.45]	[−26.88; 3.22]
prompt=emp × *favorites count*	12.42*	12.41*	8.18	8.41
	[2.17; 22.66]	[1.93; 22.55]	[−3.49; 19.86]	[−3.30; 20.12]
prompt=emp × *followers count*	2.97	3.07	4.76	4.90
	[−8.46; 14.44]	[−8.26; 14.45]	[−8.32; 17.84]	[−8.20; 18.00]
prompt=emp × *friends count*	5.23	5.24	8.30	8.44
	[−5.16; 15.60]	[−5.18; 15.80]	[−3.26; 19.87]	[−3.16; 20.04]
Num. obs.	246.00	246.00	246.00	246.00

* 0 outside the credible/confidence interval

Table S8: Results of the analysis of *Survey Question 2* using different model specifications: with vs. without a treatments interaction term, and using BRMS vs. LMER.

	Outcome: Survey Q3 (”I can understand why some people might identify with the views shown in this feed”)			
	BRMS	*BRMS* *w/ interactions*	*LMER*	*LMER* *w/ interactions*
(Intercept)	55.01^*	56.26^*	55.44^*	57.09^*
	[48.41; 61.66]	[48.03; 64.37]	[48.74; 62.15]	[48.74; 65.45]
feed=opp	-12.12^*	-13.89^*	-12.34^*	-14.66^*
	[−18.90; −5.25]	[−23.33; −4.17]	[−19.15; −5.54]	[−24.40; −4.91]
prompt=emp	9.31^*	6.88	9.80^*	6.49
	[3.10; 15.51]	[−4.56; 18.15]	[3.27; 16.33]	[−5.45; 18.43]
feed=opp × *prompt=emp*		3.43		4.54
		[−9.96; 16.69]		[−9.09; 18.17]
Num. obs.	235.00	235.00	235.00	235.00

* 0 outside the credible/confidence interval

Table S9: Results of the analysis of *Survey Question 3* using different model specifications: with vs. without a treatments interaction term, and using BRMS vs. LMER.

	Outcome: Survey Q3 (”I can understand why some people might identify with the views shown in this feed”)			
	BRMS	*BRMS* *w/ interactions*	*LMER*	*LMER* *w/ interactions*
(Intercept)	55.45*	55.80*	54.82*	55.92*
	[48.36; 62.42]	[47.24; 64.36]	[47.59; 62.04]	[47.16; 64.68]
feed=opp	−10.69*	−11.24*	−9.92*	−11.49*
	[−18.00; −3.48]	[−21.31; −1.28]	[−17.21; −2.63]	[−21.61; −1.38]
prompt=emp	7.54*	6.73	8.29*	5.99
	[1.02; 14.12]	[−5.05; 18.44]	[1.37; 15.22]	[−6.22; 18.21]
feed=opp × *prompt=emp*		1.14		3.15
		[−12.21; 14.73]		[−10.52; 16.83]
days active	4.04	3.96	2.12	2.12
	[−3.65; 11.63]	[−3.75; 11.75]	[−6.04; 10.29]	[−6.06; 10.31]
statuses count	−12.17	−12.17	−10.33	−10.17
	[−26.39; 1.94]	[−26.31; 2.17]	[−25.72; 5.06]	[−25.60; 5.25]
favorites count	7.39	7.32	7.47	7.44
	[−3.02; 17.83]	[−3.23; 17.77]	[−3.70; 18.64]	[−3.76; 18.64]
followers count	4.97	4.95	3.43	3.16
	[−4.81; 14.85]	[−4.95; 14.86]	[−7.39; 14.25]	[−7.74; 14.06]
friends count	3.87	3.98	5.85	5.89
	[−6.23; 14.01]	[−6.25; 14.07]	[−4.86; 16.56]	[−4.85; 16.63]
feed=opp × *days active*	0.89	1.02	2.26	2.33
	[−6.68; 8.42]	[−6.64; 8.65]	[−5.31; 9.84]	[−5.25; 9.91]
feed=opp × *statuses count*	−0.63	−0.60	−3.46	−3.49
	[−15.96; 14.30]	[−15.76; 14.47]	[−19.42; 12.49]	[−19.47; 12.49]
feed=opp × *favorites count*	−2.16	−2.09	−3.11	−3.17
	[−13.19; 8.91]	[−13.23; 8.97]	[−14.56; 8.34]	[−14.63; 8.29]
feed=opp × *followers count*	−3.60	−3.56	−0.95	−0.66
	[−14.80; 7.62]	[−14.90; 7.51]	[−12.65; 10.76]	[−12.42; 11.11]
feed=opp × *friends count*	−3.64	−3.77	−5.29	−5.40
	[−14.79; 7.46]	[−14.81; 7.38]	[−16.51; 5.93]	[−16.64; 5.83]
prompt=emp × *days active*	−5.84	−5.85	−5.28	−5.33
	[−12.68; 1.04]	[−12.88; 1.16]	[−13.03; 2.47]	[−13.12; 2.46]
prompt=emp × *statuses count*	5.76	5.70	8.62	8.57
	[−7.36; 18.85]	[−7.35; 18.69]	[−5.73; 22.97]	[−5.83; 22.98]
prompt=emp × *favorites count*	−2.93	−2.90	−3.51	−3.37
	[−13.08; 7.39]	[−13.11; 7.26]	[−14.73; 7.71]	[−14.66; 7.91]
prompt=emp × *followers count*	−0.29	−0.28	−3.11	−3.22
	[−11.17; 10.55]	[−11.30; 10.69]	[−15.14; 8.92]	[−15.30; 8.86]
prompt=emp × *friends count*	−3.39	−3.38	−3.87	−3.79
	[−13.58; 6.76]	[−13.39; 6.78]	[−14.95; 7.20]	[−14.91; 7.33]
Num. obs.	235.00	235.00	235.00	235.00

* 0 outside the credible/confidence interval

Table S10: Results of the analysis of *Survey Question 3* using different model specifications: with vs. without a treatments interaction term, and using BRMS vs. LMER.

	Outcome: Survey Q4 (”In the future, I would be interested in having a conversation with this feed’s owner”)			
	BRMS	*BRMS* *w/ interactions*	*LMER*	*LMER* *w/ interactions*
(Intercept)	31.33*	27.68*	26.97*	23.26*
	[23.12; 39.65]	[17.73; 37.75]	[18.84; 35.10]	[13.24; 33.28]
feed=opp	7.87	13.07*	12.51*	17.88*
	[−0.71; 16.42]	[1.34; 24.78]	[4.55; 20.46]	[6.31; 29.44]
prompt=emp	8.28*	15.34*	7.64	14.50*
	[0.41; 15.97]	[1.65; 28.72]	[−0.74; 16.01]	[0.66; 28.33]
feed=opp × *prompt=emp*		−10.20		−9.53
		[−26.17; 5.94]		[−24.71; 5.64]
Num. obs.	235.00	235.00	235.00	235.00

* 0 outside the credible/confidence interval

Table S11: Results of the analysis of *Survey Question 4* using different model specifications: with vs. without a treatments interaction term, and using BRMS vs. LMER.

	Outcome: Survey Q4 ("In the future, I would be interested in having a conversation with this feed's owner")			
	BRMS	*BRMS* *w/ interactions*	*LMER*	*LMER* *w/ interactions*
(Intercept)	28.90*	24.36*	27.02*	22.40*
	[20.88; 36.98]	[14.88; 33.84]	[18.83; 35.20]	[12.64; 32.16]
feed=opp	9.89*	16.39*	11.60*	18.30*
	[1.70; 18.04]	[5.27; 27.38]	[3.45; 19.75]	[7.07; 29.53]
prompt=emp	6.68	15.88*	7.58	16.91*
	[−0.85; 14.17]	[2.88; 28.81]	[−0.42; 15.59]	[3.48; 30.34]
feed=opp × *prompt=emp*		−12.95		−12.96
		[−27.80; 2.26]		[−27.94; 2.01]
days active	−11.50*	−11.82*	−11.53*	−11.46*
	[−20.57; −2.49]	[−20.63; −2.81]	[−20.66; −2.39]	[−20.55; −2.38]
statuses count	4.17	3.89	1.69	1.01
	[−12.10; 20.44]	[−12.28; 19.79]	[−15.84; 19.21]	[−16.44; 18.46]
favorites count	2.21	2.13	1.07	1.00
	[−9.48; 13.88]	[−9.44; 14.00]	[−11.09; 13.24]	[−11.10; 13.10]
followers count	8.05	9.62	11.63	12.58
	[−6.14; 22.18]	[−4.30; 23.58]	[−3.20; 26.47]	[−2.24; 27.40]
friends count	−3.22	−3.37	−3.47	−3.31
	[−15.74; 9.44]	[−15.82; 9.09]	[−16.22; 9.29]	[−16.00; 9.38]
feed=opp × *days active*	1.96	1.98	4.37	3.79
	[−6.81; 10.91]	[−6.84; 10.72]	[−4.03; 12.77]	[−4.61; 12.19]
feed=opp × *statuses count*	−0.17	−0.48	0.34	0.54
	[−17.28; 16.52]	[−16.92; 16.51]	[−17.14; 17.82]	[−16.86; 17.94]
feed=opp × *favorites count*	−0.26	−0.13	0.75	1.25
	[−12.86; 12.57]	[−12.60; 12.22]	[−11.33; 12.83]	[−10.80; 13.30]
feed=opp × *followers count*	−15.63*	−17.30*	−17.43*	−18.44*
	[−30.39; −0.94]	[−32.02; −2.31]	[−32.15; −2.72]	[−33.17; −3.71]
feed=opp × *friends count*	5.65	6.24	3.60	3.83
	[−7.46; 18.98]	[−6.87; 19.36]	[−9.34; 16.55]	[−9.08; 16.73]
prompt=emp × *days active*	0.24	0.50	−1.50	−1.32
	[−7.99; 8.42]	[−7.49; 8.54]	[−10.65; 7.65]	[−10.40; 7.77]
prompt=emp × *statuses count*	−13.04	−12.67	−10.04	−9.35
	[−27.84; 2.13]	[−27.27; 1.93]	[−26.61; 6.53]	[−25.82; 7.13]
prompt=emp × *favorites count*	12.72*	12.34*	11.64	11.24
	[1.20; 24.27]	[0.76; 23.84]	[−1.45; 24.73]	[−1.76; 24.24]
prompt=emp × *followers count*	6.30	6.75	4.39	4.66
	[−6.45; 18.85]	[−6.06; 19.56]	[−10.06; 18.84]	[−9.69; 19.02]
prompt=emp × *friends count*	−4.16	−4.63	−3.26	−3.87
	[−15.83; 7.46]	[−16.36; 7.06]	[−16.15; 9.63]	[−16.69; 8.94]
Num. obs.	235.00	235.00	235.00	235.00

* 0 outside the credible/confidence interval

Table S12: Results of the analysis of *Survey Question 4* using different model specifications: with vs. without a treatments interaction term, and using BRMS vs. LMER.

S3 Covariate Balance

In this section, we test whether the covariates are similar on average across the different treatment arms. We use two techniques: (*i*) regression analysis, regressing the treatment assignments on the covariates, and (*ii*) permutation tests, comparing the goodness of the fit of the empirical treatment assignment with many other simulated treatment assignments.

In both analyses, we consider five covariates (days active, number of statuses, number of favorites, number of friends, number of followers) that capture the users' tenure on Twitter, their level of activity, influences, and overall behavior. (We consider the same set of covariates in our main analyses.) We note that while many other covariates would be useful to include in our analyses, we were limited by the information we could collect about the users that participated in the experiment.

S3.1 Regression Analyses

We start by regressing the treatment assignments on the covariates. In Table S13, we report the results for the feed treatment assignments (opposite vs. same), and in Table S14, we report the results for the prompt treatment assignment (empathic vs. control). In both cases, we run five regressions, one considering all users and one for each of the four survey questions, considering only the users who responded to that question. Since the feed treatment assignments are on the session level, i.e., one user can be assigned both same and opposite feeds in different sessions, we include user and giver fixed effects. In Tables S13 and S14, we report the results of Bayesian regressions using *brms* (family=bernoulli, coefficient priors: $\mathcal{N}(0, 30)$, standard deviation priors: $\mathcal{N}(0, 1)$). The corresponding *glmer* models, led to the same substantive results.

In both treatments, we do not find any significant differences in the background covariates between the users in different treatment arms. The only statistically significant terms are the intercept terms for the feed treatment assignment in survey questions 1 through 4. The positive intercepts suggest that users who saw opposite feeds were more likely to respond to the survey.

S3.2 Randomization Inference

Next, we use randomization inference to test whether the observed covariate imbalances are larger than we would expect from chance alone. First, we regress the observed treatment assignment on the covariates and record the log-likelihood of the model. Then, we repeat this analysis 100,000 times using different treatment reassignments to obtain the null distribution. We reassign the treatments using the same randomization scheme as in the actual experiment, i.e., assigning an empathic vs. control prompt completely at random and assigning a feed from one of the 200 feed givers completely at random. Using the observed log-likelihood and the null distribution, we compute a two-sided p-value (computing both one-sided p-values and doubling the smaller one [5]). Since fitting the Baysian models takes significantly longer, too long for us to repeat the analyses many times, we used their frequentist equivalents instead (*glm* and *glmer*). We perform this analysis for both the feed and prompt treatments, considering all users and only the users that responded to a given survey question (Q1-Q4).

Figures S3 and S4 show the observed log-likelihood (red line), the null distributions, and the p-values for the feed and prompt treatment assignments. We find that in all cases, the p-values are much larger than the conventional threshold of 0.05.

	Outcome: Feed Assignment				
	All	*Survey Q1*	*Survey Q2*	*Survey Q3*	*Survey Q4*
(Intercept)	−0.00	1.69*	1.77*	2.03*	1.86*
	[−0.38; 0.37]	[0.75; 2.76]	[0.80; 2.88]	[1.09; 3.14]	[0.93; 2.94]
days active	−0.04	−0.15	0.13	0.00	−0.00
	[−0.14; 0.06]	[−0.87; 0.54]	[−0.59; 0.88]	[−0.73; 0.74]	[−0.73; 0.73]
statuses count	−0.01	0.47	0.38	0.45	0.58
	[−0.17; 0.16]	[−0.78; 1.75]	[−0.92; 1.70]	[−0.92; 1.80]	[−0.73; 1.93]
favorites count	0.01	−0.19	−0.23	−0.27	−0.16
	[−0.11; 0.13]	[−1.21; 0.83]	[−1.32; 0.84]	[−1.37; 0.84]	[−1.28; 0.94]
followers count	−0.12	0.03	−0.03	0.04	0.07
	[−0.27; 0.02]	[−1.05; 1.11]	[−1.12; 1.11]	[−1.06; 1.16]	[−1.13; 1.31]
friends count	0.12	−0.04	−0.28	−0.01	−0.17
	[−0.01; 0.24]	[−1.01; 0.93]	[−1.34; 0.73]	[−1.08; 1.03]	[−1.25; 0.87]

* 0 outside the credible interval

Table S13: Results of regressing the *feed* treatment assignments (same vs. opposite) on the user covariates, testing for covariate imbalance. The model includes user and feed giver fixed effects. The different models consider all users and the subset of users that responded to each of the four survey questions.

	Outcome: Prompt Assignment				
	All	*Survey Q1*	*Survey Q2*	*Survey Q3*	*Survey Q4*
(Intercept)	−0.03	0.04	−0.00	0.01	0.01
	[−0.14; 0.07]	[−0.25; 0.33]	[−0.30; 0.29]	[−0.30; 0.31]	[−0.30; 0.31]
days active	−0.08	−0.05	−0.08	−0.10	−0.10
	[−0.20; 0.04]	[−0.39; 0.30]	[−0.42; 0.24]	[−0.46; 0.25]	[−0.46; 0.25]
statuses count	−0.01	−0.16	−0.05	−0.17	−0.17
	[−0.20; 0.18]	[−0.77; 0.44]	[−0.67; 0.56]	[−0.79; 0.46]	[−0.79; 0.46]
favorites count	−0.07	−0.09	−0.17	−0.22	−0.22
	[−0.21; 0.08]	[−0.55; 0.38]	[−0.64; 0.29]	[−0.72; 0.27]	[−0.72; 0.27]
followers count	0.01	−0.36	−0.38	−0.26	−0.26
	[−0.16; 0.18]	[−0.88; 0.15]	[−0.91; 0.14]	[−0.80; 0.26]	[−0.80; 0.26]
friends count	0.02	0.30	0.34	0.42	0.42
	[−0.13; 0.16]	[−0.17; 0.77]	[−0.12; 0.81]	[−0.08; 0.93]	[−0.08; 0.93]

* 0 outside the credible interval

Table S14: Results of regressing the *prompt* treatment assignments (empathic vs. control) on the user covariates, testing for covariate imbalance. The different models consider all users and the subset of users that responded to each of the four survey questions.

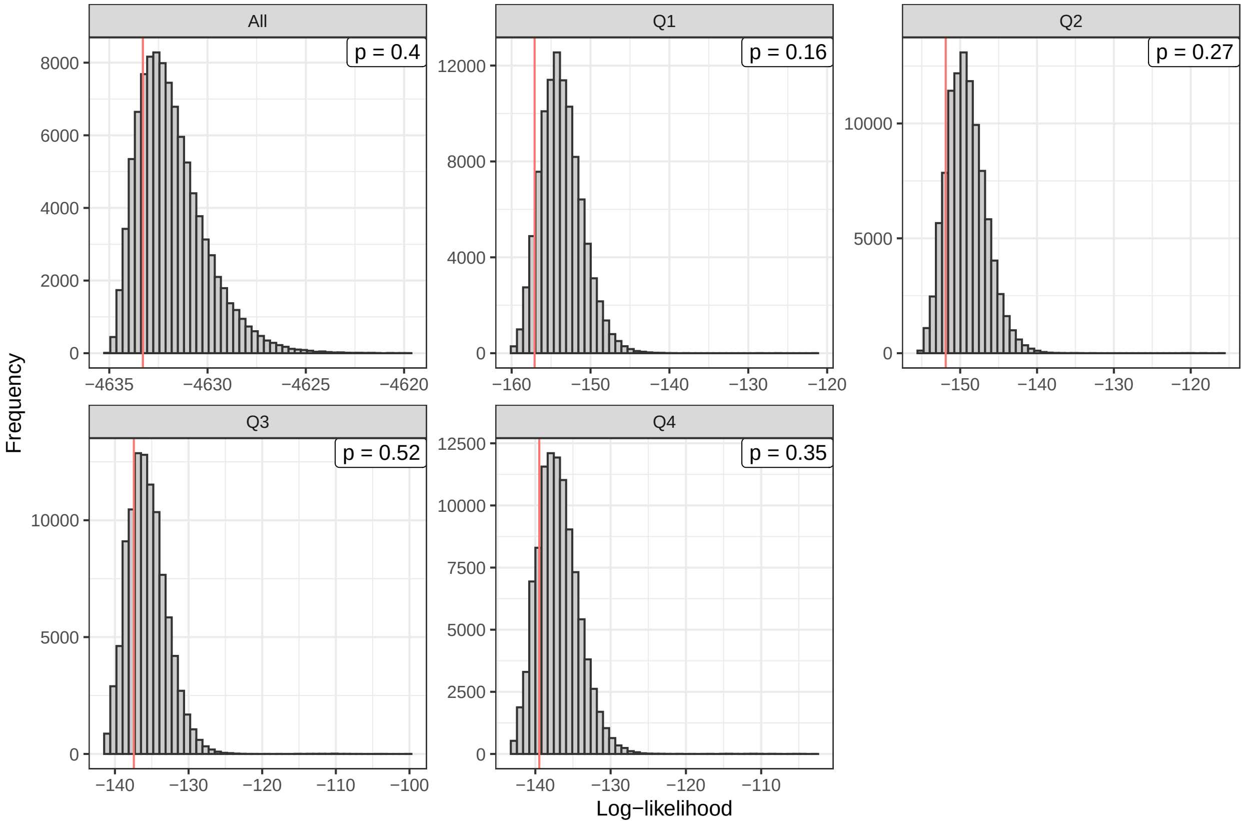

Figure S3: Randomization inference tests for covariate imbalance of the *feed* treatment assignment. The histogram shows the null distribution of the log-likelihoods obtained by reassigning treatments using the same randomization scheme as in the actual experiment. The red line shows the log-likelihood of the actual treatment assignment. The two-sided p-value is shown in the top-right corner.

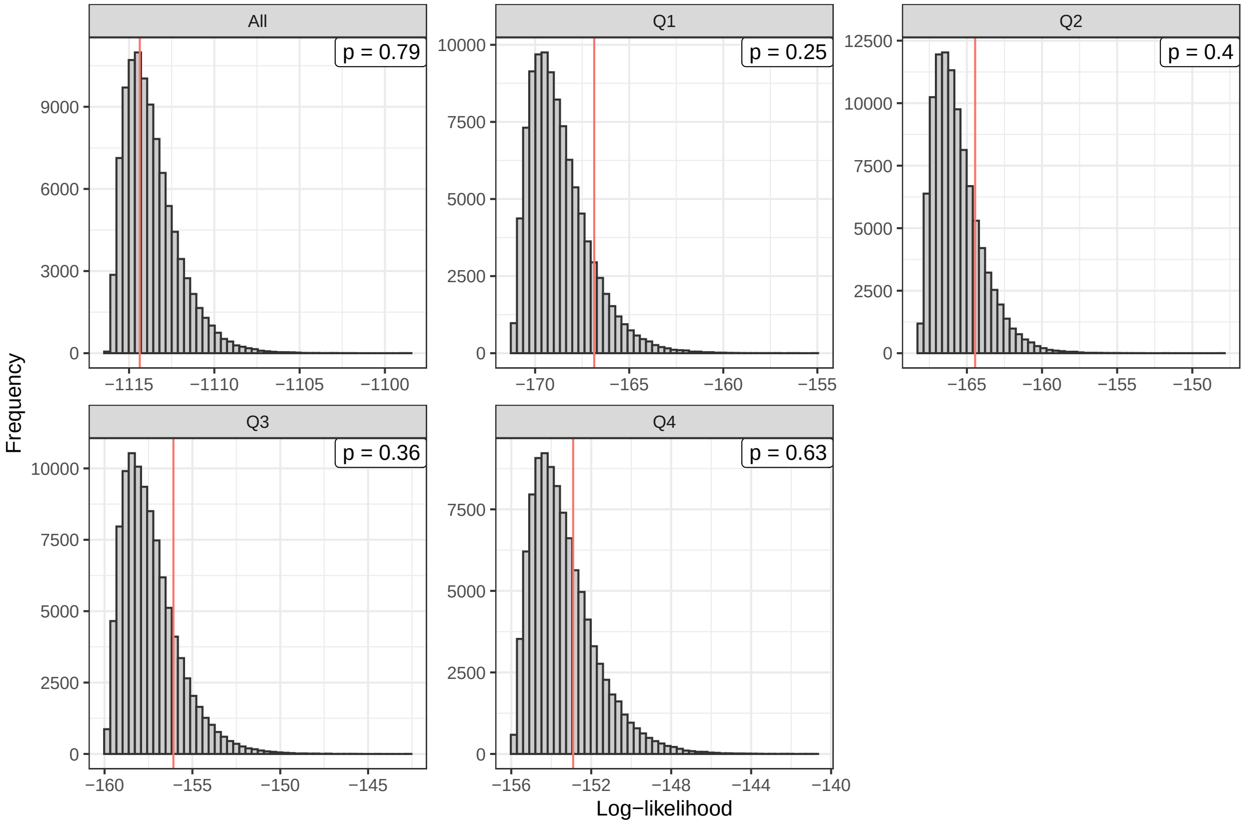

Figure S4: Randomization inference tests for covariate imbalance of the *prompt* treatment assignment. The histogram shows the null distribution of the log-likelihoods obtained by reassigning treatments using the same randomization scheme as in the actual experiment. The red line shows the log-likelihood of the actual treatment assignment. The two-sided p-value is shown in the top-right corner.

S4 Attrition Analyses

Following the guidance provided in [6], we run simulations to better-understand how selective attrition in survey responses between treated and untreated participants might affect our average treatment effects (ATEs). Table S15 provides descriptive statistics for covariate and outcomes data available across sessions that are needed in order to compute treatment effects. The table shows that only a small fraction of total sessions were shown a survey, and a subset of those provided answers to the different survey questions.

	All	With ideology	Survey shown	Q1	Q2	Q3	Q4
# sessions	10632	6687	405	253	246	235	235
# sessions feed=same	3366	3366	135	81	78	66	68
# sessions feed=opp	3321	3321	270	172	168	169	167
# sessions feed=unknown	3945	0	0	0	0	0	0
# sessions prompt=empathic	5431	3507	208	133	128	121	121
# sessions prompt=control	5201	3180	197	120	118	114	114
# participants	2592	1611	290	200	196	184	188
# participants prompt=empathic	1277	786	146	100	96	90	91
# participants prompt=control	1315	825	144	100	100	94	97
# feed "givers"	200	200	165	133	132	128	127

Table S15: Descriptive statistics for different slices of data. *All*: all data collected; *With ideology*: sample of sessions for which we can infer the users' political alignment; *Survey shown*: sessions where users were shown a survey; *Q1-Q4*: sessions where the users responded to the questions.

As the authors of [6] explain, in certain attrition scenarios, outcome data might be missing independent of potential outcomes (MIPO), i.e., learning whether there is missing outcome data reveals no information about what those outcomes could have been for a particular participant. This requires no adjustment to the observed outcome data when measuring treatment effects. A relaxation of this scenario is MIPO | X: conditional on a basket of covariates X, missing data is independent of potential outcomes. In this scenario, outcomes can be re-weighted according to any covariate imbalances that correlate with missingness. In our case, we see that attrition in survey response differs by treatment value for each of our treatments; hence, invoking MIPO or MIPO | X is likely not sufficient to correct for potential biases in our outcomes introduced by selective attrition.

One potential path forward involves identifying bounds that reflect the range of plausible estimates for each ATE on survey responses. A conservative approach is to use "extreme value bounds", i.e., to fill in missing values in treated and control groups with the highest and lowest possible values (or vice versa) and then compute the ATE under each scenario to produce upper and lower bounds, respectively. Unfortunately, this method often produces bounds that are too wide to be useful. Another method, monotonicity bounds, makes an additional assumption in order to bracket the true ATE in a more realistic fashion. In particular, it assumes that attrition is monotonic, i.e., that any participant whose outcome data would not be missing if assigned to the treatment group also would not be missing if assigned to the control group (or vice versa—participants with outcomes that would not be missing if assigned to the control group also would not be missing if assigned to treatment). The method then involves trimming outcomes observed from members of the treatment group that are unlikely to report if they belonged in the control group (or vice versa) and using the trimmed set of observations in order to estimate upper and lower bounds for each ATE (see [6] for more details on this method).

We produce monotonicity bounds for the prompt and feed ATEs reported in the main text. Figure S5 shows these bounds for the feed treatment, and Figure S6 for the prompt treatment, for each of the four survey questions. Both figures are produced using Bayesian regression models with a specification similar to the models used to compute the main effects. Bounds produced using linear mixed effects models are very similar; hence, we do not include them below.

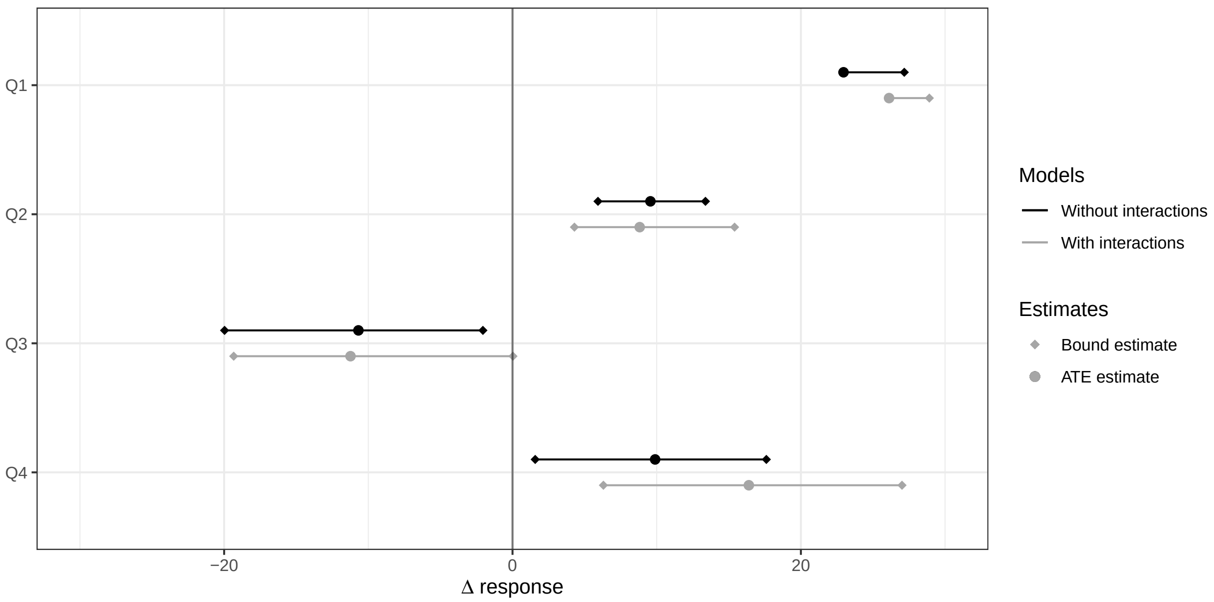

Figure S5: Monotonicity bounds for the **feed treatment** ATE. Black lines indicate bounds produced for models without interaction effects between both treatments; gray includes interaction effects. Circles depict the ATE point estimate produced in the main analyses.

We can make several observations from the figures. For one, there are many questions where the ATE point estimate coincides with, or is very close to, the lower bound for a given outcome. This is because attrition is not very different between treatment and control, and hence, very few observations (if any) are “trimmed” from the analysis in order to satisfy the monotonicity assumption. We can also see that in all cases but two (for models that include interaction effects), the bounds do not intersect zero. This suggests that, under the assumption of monotonicity, attrition does not appear to have significantly affected the sign of our point estimates. Tables S16, S17, S18, S19 and S20, S21, S22, S23 offer additional details on the upper and lower bounds produced by the Bayesian regression models for the feed and prompt treatments, respectively.

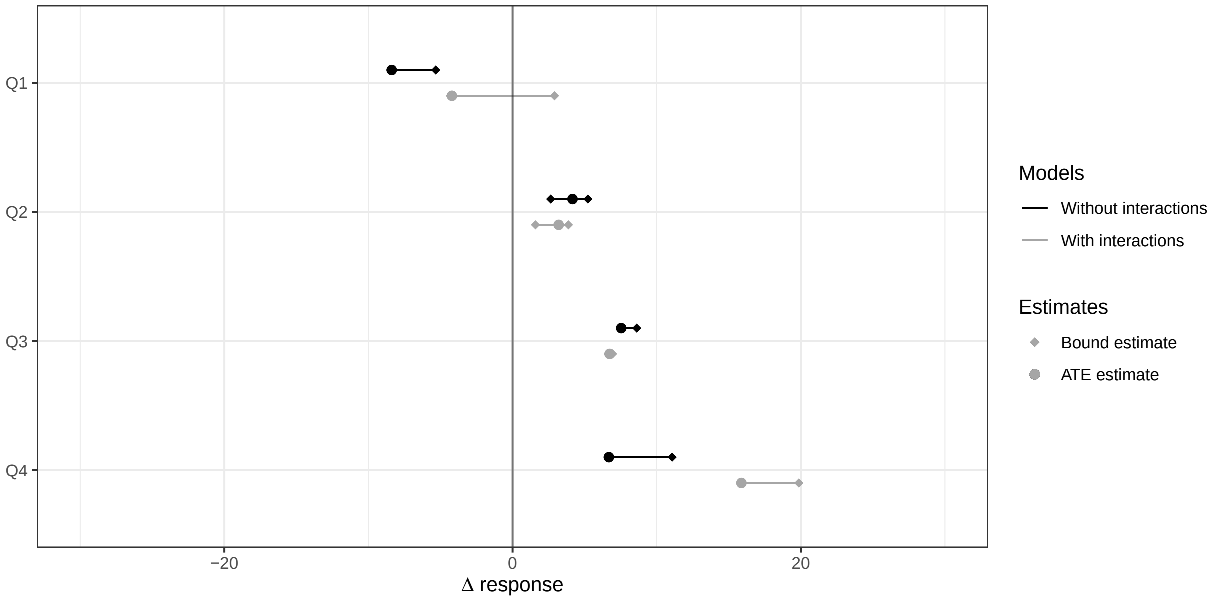

Figure S6: Monotonicity bounds for the **prompt treatment** ATE. Black lines indicate bounds produced for models without interaction effects between both treatments; gray includes interaction effects. Circles depict the ATE point estimate produced in the main analyses.

	Treatment: Feed, Outcome: Survey Q1			
	(”This feed is different from what I’m used to seeing”)			
	lower bound	*upper bound*	*lower bound w/ interactions*	*upper bound w/ interactions*
(Intercept)	60.19*	58.53*	58.15*	57.37*
	[53.19; 67.10]	[52.38; 64.74]	[49.93; 66.42]	[50.02; 64.72]
feed=opp	23.00*	27.17*	26.07*	28.91*
	[15.90; 30.04]	[20.78; 33.54]	[16.23; 35.76]	[20.21; 37.67]
prompt=emp	−8.40*	−6.36*	−4.32	−4.10
	[−15.12; −1.72]	[−12.36; −0.36]	[−15.55; 7.14]	[−14.17; 5.99]
feed=opp × *prompt=emp*			−5.96	−3.33
			[−19.19; 7.29]	[−15.41; 8.57]
days active	−4.94	−4.89	−5.16	−5.02
	[−12.90; 3.02]	[−11.89; 2.20]	[−12.92; 2.76]	[−11.99; 1.99]
statuses count	−5.66	−3.96	−5.58	−3.73
	[−19.67; 8.19]	[−16.41; 8.57]	[−19.45; 8.32]	[−16.14; 8.66]
favorites count	−5.66	−6.23	−5.83	−6.38
	[−15.96; 4.69]	[−15.51; 2.93]	[−16.05; 4.38]	[−15.42; 2.71]
followers count	2.55	4.25	3.27	4.57
	[−7.15; 12.45]	[−4.46; 13.06]	[−6.85; 13.28]	[−4.11; 13.22]
friends count	−7.82	−8.09	−7.72	−8.09
	[−17.61; 2.01]	[−16.76; 0.49]	[−17.64; 2.02]	[−16.72; 0.58]
feed=opp × *days active*	2.59	3.58	2.62	3.66
	[−5.33; 10.35]	[−3.32; 10.51]	[−5.14; 10.38]	[−3.34; 10.51]
feed=opp × *statuses count*	−1.15	1.44	−1.57	1.04
	[−15.59; 13.38]	[−11.69; 14.62]	[−16.13; 12.91]	[−12.03; 14.06]
feed=opp × *favorites count*	11.93*	10.95*	12.14*	11.12*
	[0.99; 22.95]	[0.98; 20.85]	[1.16; 23.25]	[1.38; 20.90]
feed=opp × *followers count*	1.17	−0.97	0.47	−1.29
	[−10.24; 12.39]	[−11.21; 9.17]	[−11.12; 12.07]	[−11.42; 8.83]
feed=opp × *friends count*	5.99	4.21	6.11	4.36
	[−5.23; 17.19]	[−5.52; 13.90]	[−4.88; 17.18]	[−5.45; 14.27]
prompt=emp × *days active*	5.69	5.51	5.87	5.58
	[−1.67; 12.92]	[−1.10; 12.09]	[−1.33; 13.09]	[−0.81; 12.04]
prompt=emp × *statuses count*	1.75	1.09	1.90	1.18
	[−11.53; 14.94]	[−11.02; 13.25]	[−11.40; 15.22]	[−11.07; 13.33]
prompt=emp × *favorites count*	2.28	1.78	2.12	1.65
	[−8.12; 12.67]	[−7.88; 11.34]	[−8.14; 12.41]	[−7.82; 11.18]
prompt=emp × *followers count*	−0.03	−4.84	0.11	−4.73
	[−11.28; 11.29]	[−15.34; 5.61]	[−11.23; 11.60]	[−15.09; 5.51]
prompt=emp × *friends count*	1.00	2.18	0.72	2.02
	[−9.55; 11.42]	[−7.40; 11.70]	[−9.79; 11.27]	[−7.55; 11.46]
Num. obs.	253.00	243.00	253.00	243.00

* 0 outside the credible interval.

Table S16: BRMS monotonicity estimates for feed ATE for *Survey Question 1*.

	Treatment: Feed, Outcome: Survey Q2 (”I learned something new from browsing this person’s feed”)			
	lower bound	*upper bound*	*lower bound* *w/ interactions*	*upper bound* *w/ interactions*
(Intercept)	33.76*	35.61*	34.81*	34.31*
	[27.48; 39.92]	[29.21; 42.04]	[27.32; 42.34]	[26.79; 41.99]
feed=opp	5.93	13.38*	4.29	15.40*
	[−0.49; 12.40]	[6.73; 20.03]	[−4.92; 13.46]	[5.86; 24.62]
prompt=emp	5.11	0.77	2.93	3.34
	[−1.04; 11.29]	[−5.44; 7.04]	[−7.58; 13.34]	[−7.36; 13.68]
feed=opp × *prompt=emp*			3.28	−3.87
			[−9.17; 15.99]	[−16.29; 8.94]
days active	−6.31	−7.03	−6.08	−7.24
	[−13.41; 0.86]	[−14.29; 0.34]	[−13.36; 1.11]	[−14.49; 0.05]
statuses count	9.48	11.42	9.17	11.56
	[−3.47; 22.32]	[−1.78; 24.62]	[−3.81; 22.22]	[−1.62; 24.74]
favorites count	−3.85	−3.28	−3.57	−3.44
	[−13.06; 5.61]	[−12.64; 6.25]	[−12.94; 5.75]	[−13.05; 6.00]
followers count	1.54	1.34	1.21	1.75
	[−7.43; 10.60]	[−7.86; 10.45]	[−7.99; 10.29]	[−7.58; 10.99]
friends count	−10.72*	−11.41*	−10.73*	−11.29*
	[−19.87; −1.49]	[−20.64; −2.15]	[−19.79; −1.54]	[−20.61; −1.89]
feed=opp × *days active*	6.48	4.88	6.46	4.91
	[−0.73; 13.57]	[−2.49; 12.19]	[−0.67; 13.66]	[−2.32; 12.19]
feed=opp × *statuses count*	−18.81*	−15.22*	−18.36*	−15.45*
	[−32.36; −4.99]	[−29.07; −1.33]	[−32.16; −4.60]	[−29.68; −1.40]
feed=opp × *favorites count*	−0.23	0.85	−0.50	1.02
	[−10.45; 9.94]	[−9.49; 11.14]	[−10.59; 9.70]	[−9.33; 11.38]
feed=opp × *followers count*	1.18	3.89	1.49	3.39
	[−9.28; 11.71]	[−7.10; 14.76]	[−9.19; 12.01]	[−7.84; 14.62]
feed=opp × *friends count*	9.36	9.43	9.18	9.60
	[−0.83; 19.50]	[−0.94; 19.91]	[−1.09; 19.42]	[−1.00; 20.19]
prompt=emp × *days active*	−0.28	1.20	−0.50	1.37
	[−7.03; 6.26]	[−5.82; 8.18]	[−7.37; 6.23]	[−5.66; 8.30]
prompt=emp × *statuses count*	−10.14	−11.78	−10.24	−11.77
	[−22.83; 2.44]	[−24.73; 1.12]	[−22.82; 2.61]	[−24.56; 1.11]
prompt=emp × *favorites count*	12.91*	11.24*	12.99*	11.12*
	[3.27; 22.58]	[1.21; 21.15]	[3.29; 22.73]	[1.10; 21.24]
prompt=emp × *followers count*	4.69	0.99	4.68	1.14
	[−5.85; 15.25]	[−10.25; 12.26]	[−6.00; 15.20]	[−10.19; 12.55]
prompt=emp × *friends count*	3.55	6.01	3.75	5.70
	[−6.10; 13.22]	[−4.05; 16.17]	[−6.20; 13.54]	[−4.28; 15.79]
Num. obs.	234.00	230.00	234.00	230.00

[*] 0 outside the credible interval.

Table S17: BRMS monotonicity estimates for feed ATE for *Survey Question 2*.

	Treatment: Feed, Outcome: Survey Q3 (”I can understand why some people might identify with the views shown in this feed”)			
	lower bound	*upper bound*	*lower bound w/ interactions*	*upper bound w/ interactions*
(Intercept)	56.18*	56.44*	55.74*	55.07*
	[50.32; 61.97]	[50.58; 62.29]	[48.58; 62.90]	[48.12; 61.99]
feed=opp	−19.97*	−2.05	−19.34*	0.03
	[−26.16; −13.80]	[−8.19; 4.11]	[−27.83; −10.70]	[−8.44; 8.67]
prompt=emp	6.62*	5.80*	7.45	8.64
	[0.76; 12.53]	[0.12; 11.56]	[−2.37; 17.26]	[−1.05; 18.32]
feed=opp × *prompt=emp*			−1.22	−4.16
			[−12.90; 10.40]	[−15.82; 7.38]
days active	2.18	1.16	2.16	1.10
	[−4.40; 8.76]	[−5.42; 7.70]	[−4.50; 8.82]	[−5.51; 7.65]
statuses count	−10.47	−8.18	−10.40	−8.23
	[−22.50; 1.67]	[−20.13; 3.88]	[−22.53; 1.82]	[−20.19; 3.72]
favorites count	6.06	7.41	6.04	7.47
	[−2.93; 15.13]	[−1.27; 16.14]	[−2.99; 15.07]	[−1.35; 16.24]
followers count	4.05	3.98	4.19	4.35
	[−4.26; 12.30]	[−4.28; 12.23]	[−4.09; 12.52]	[−3.92; 12.56]
friends count	4.27	3.11	4.27	3.03
	[−4.38; 12.92]	[−5.31; 11.86]	[−4.42; 12.95]	[−5.47; 11.46]
feed=opp × *days active*	−0.13	2.19	−0.09	2.05
	[−6.55; 6.43]	[−4.33; 8.75]	[−6.80; 6.44]	[−4.49; 8.60]
feed=opp × *statuses count*	−0.13	6.83	−0.33	6.79
	[−13.60; 13.24]	[−6.08; 20.04]	[−13.87; 13.18]	[−6.07; 19.83]
feed=opp × *favorites count*	−4.10	−2.68	−3.97	−2.64
	[−13.74; 5.73]	[−12.20; 6.73]	[−13.78; 5.86]	[−12.11; 6.83]
feed=opp × *followers count*	−1.25	−6.85	−1.39	−7.35
	[−11.40; 8.75]	[−16.61; 3.06]	[−11.26; 8.57]	[−16.99; 2.39]
feed=opp × *friends count*	−1.75	−3.88	−1.71	−3.69
	[−11.34; 7.73]	[−13.43; 5.70]	[−11.29; 7.92]	[−13.27; 5.96]
prompt=emp × *days active*	−3.08	−1.38	−3.06	−1.25
	[−9.36; 3.14]	[−7.75; 4.93]	[−9.36; 3.20]	[−7.68; 5.08]
prompt=emp × *statuses count*	−1.60	−1.60	−1.58	−1.61
	[−14.59; 11.48]	[−13.28; 9.92]	[−14.54; 11.41]	[−13.41; 10.13]
prompt=emp × *favorites count*	1.63	−3.24	1.54	−3.45
	[−8.10; 11.13]	[−12.01; 5.71]	[−8.07; 11.12]	[−12.39; 5.55]
prompt=emp × *followers count*	4.40	1.11	4.42	1.41
	[−6.20; 14.86]	[−8.68; 10.99]	[−6.06; 15.01]	[−8.63; 11.41]
prompt=emp × *friends count*	−4.40	−2.11	−4.43	−2.32
	[−13.36; 4.54]	[−11.53; 7.17]	[−13.58; 4.63]	[−11.44; 6.93]
Num. obs.	198.00	198.00	198.00	198.00

* 0 outside the credible interval.

Table S18: BRMS monotonicity estimates for feed ATE for *Survey Question 3*.

	Treatment: Feed, Outcome: Survey Q4			
	(”In the future, I would be interested in having a conversation with this feed’s owner”)			
	lower bound	*upper bound*	*lower bound w/ interactions*	*upper bound w/ interactions*
(Intercept)	27.81*	30.26*	24.73*	24.26*
	[21.20; 34.54]	[22.83; 37.67]	[16.76; 32.86]	[15.65; 32.84]
feed=opp	1.57	17.60*	6.30	27.01*
	[−5.55; 8.71]	[9.88; 25.27]	[−3.76; 16.04]	[16.65; 37.53]
prompt=emp	8.90*	4.49	15.09*	16.37*
	[2.41; 15.31]	[−2.92; 11.89]	[4.05; 26.13]	[4.75; 28.09]
feed=opp × *prompt=emp*			−9.21	−18.05*
			[−22.74; 4.23]	[−32.08; −4.21]
days active	−7.83	−13.48*	−8.25*	−13.85*
	[−15.80; 0.19]	[−22.07; −4.87]	[−16.25; −0.36]	[−22.07; −5.58]
statuses count	−0.17	4.88	−0.25	4.32
	[−14.34; 14.19]	[−10.39; 20.01]	[−14.29; 13.84]	[−10.50; 19.29]
favorites count	1.61	2.14	1.75	1.98
	[−8.50; 11.70]	[−8.50; 13.06]	[−8.32; 11.88]	[−8.49; 12.68]
followers count	11.14	7.96	11.99	10.22
	[−1.09; 23.54]	[−5.45; 21.11]	[−0.37; 24.40]	[−2.98; 23.33]
friends count	−4.59	−3.99	−4.72	−4.36
	[−15.39; 6.20]	[−15.66; 7.79]	[−15.48; 6.14]	[−15.86; 7.22]
feed=opp × *days active*	6.23	2.35	6.05	2.13
	[−1.58; 14.07]	[−6.13; 10.73]	[−1.78; 13.86]	[−6.10; 10.35]
feed=opp × *statuses count*	0.41	2.63	0.32	1.83
	[−14.42; 15.20]	[−13.50; 18.50]	[−14.54; 15.08]	[−14.11; 17.78]
feed=opp × *favorites count*	−5.24	−3.69	−5.23	−3.49
	[−16.20; 5.87]	[−15.45; 8.00]	[−16.09; 5.61]	[−15.11; 7.88]
feed=opp × *followers count*	−14.88*	−12.54	−16.09*	−14.54*
	[−27.93; −1.97]	[−27.14; 2.07]	[−29.34; −2.95]	[−28.74; −0.39]
feed=opp × *friends count*	6.12	4.05	6.72	5.00
	[−5.84; 17.98]	[−8.61; 16.72]	[−4.96; 18.37]	[−7.82; 17.57]
prompt=emp × *days active*	−5.18	3.28	−4.72	3.60
	[−12.68; 2.19]	[−4.81; 11.27]	[−12.10; 2.69]	[−4.21; 11.27]
prompt=emp × *statuses count*	−7.72	−13.26	−7.36	−12.50
	[−21.55; 5.94]	[−28.33; 1.80]	[−21.12; 6.56]	[−27.16; 2.41]
prompt=emp × *favorites count*	13.81*	12.40*	13.24*	12.22*
	[3.34; 24.48]	[1.19; 23.55]	[2.76; 23.88]	[1.17; 23.15]
prompt=emp × *followers count*	2.77	5.82	3.32	5.85
	[−9.65; 14.92]	[−7.95; 19.45]	[−8.71; 15.28]	[−7.46; 19.03]
prompt=emp × *friends count*	−1.65	−2.22	−2.10	−2.76
	[−12.17; 8.90]	[−14.22; 9.77]	[−12.57; 8.50]	[−14.61; 9.22]
Num. obs.	206.00	204.00	206.00	204.00

* 0 outside the credible interval.

Table S19: BRMS monotonicity estimates for feed ATE for *Survey Question 4*.

	Treatment: Prompt, Outcome: Survey Q1 ("This feed is different from what I'm used to seeing")			
	lower bound	*upper bound*	*lower bound* *w/ interactions*	*upper bound* *w/ interactions*
(Intercept)	60.19*	61.64*	58.15*	57.78*
	[53.19; 67.10]	[54.91; 68.30]	[49.93; 66.42]	[49.90; 65.64]
feed=opp	23.00*	20.84*	26.07*	26.56*
	[15.90; 30.04]	[13.94; 27.74]	[16.23; 35.76]	[17.02; 35.87]
prompt=emp	−8.40*	−5.33	−4.32	2.91
	[−15.12; −1.72]	[−11.74; 1.07]	[−15.55; 7.14]	[−8.25; 14.09]
feed=opp × *prompt=emp*			−5.96	−11.65
			[−19.19; 7.29]	[−24.57; 1.43]
days active	−4.94	−3.58	−5.16	−3.82
	[−12.90; 3.02]	[−11.03; 3.89]	[−12.92; 2.76]	[−11.45; 3.72]
statuses count	−5.66	−5.20	−5.58	−4.80
	[−19.67; 8.19]	[−18.60; 8.05]	[−19.45; 8.32]	[−18.30; 8.60]
favorites count	−5.66	−5.18	−5.83	−5.46
	[−15.96; 4.69]	[−14.88; 4.52]	[−16.05; 4.38]	[−15.30; 4.37]
followers count	2.55	1.54	3.27	2.74
	[−7.15; 12.45]	[−8.03; 10.94]	[−6.85; 13.28]	[−6.69; 12.14]
friends count	−7.82	−6.42	−7.72	−6.23
	[−17.61; 2.01]	[−15.63; 2.77]	[−17.64; 2.02]	[−15.51; 3.07]
feed=opp × *days active*	2.59	0.63	2.62	0.61
	[−5.33; 10.35]	[−6.82; 8.12]	[−5.14; 10.38]	[−6.93; 8.18]
feed=opp × *statuses count*	−1.15	−1.86	−1.57	−2.85
	[−15.59; 13.38]	[−15.62; 11.98]	[−16.13; 12.91]	[−17.02; 11.41]
feed=opp × *favorites count*	11.93*	11.51*	12.14*	11.82*
	[0.99; 22.95]	[1.00; 21.90]	[1.16; 23.25]	[1.21; 22.40]
feed=opp × *followers count*	1.17	2.54	0.47	1.25
	[−10.24; 12.39]	[−8.51; 13.57]	[−11.12; 12.07]	[−9.72; 12.11]
feed=opp × *friends count*	5.99	3.54	6.11	3.82
	[−5.23; 17.19]	[−6.96; 13.93]	[−4.88; 17.18]	[−6.77; 14.29]
prompt=emp × *days active*	5.69	6.81*	5.87	7.08*
	[−1.67; 12.92]	[0.00; 13.64]	[−1.33; 13.09]	[0.14; 14.08]
prompt=emp × *statuses count*	1.75	1.01	1.90	1.41
	[−11.53; 14.94]	[−11.52; 13.58]	[−11.40; 15.22]	[−11.35; 14.12]
prompt=emp × *favorites count*	2.28	4.06	2.12	3.68
	[−8.12; 12.67]	[−5.80; 13.90]	[−8.14; 12.41]	[−6.41; 13.56]
prompt=emp × *followers count*	−0.03	−0.50	0.11	−0.23
	[−11.28; 11.29]	[−11.30; 10.38]	[−11.23; 11.60]	[−11.18; 10.90]
prompt=emp × *friends count*	1.00	1.96	0.72	1.54
	[−9.55; 11.42]	[−8.16; 12.01]	[−9.79; 11.27]	[−8.57; 11.63]
Num. obs.	253.00	246.00	253.00	246.00

* 0 outside the credible interval.

Table S20: BRMS monotonicity estimates for prompt ATE for *Survey Question 1*.

	Treatment: Prompt, Outcome: Survey Q2 (”I learned something new from browsing this person’s feed”)			
	lower bound	*upper bound*	*lower bound* *w/ interactions*	*upper bound* *w/ interactions*
(Intercept)	34.14*	34.09*	34.67*	34.75*
	[27.53; 40.72]	[27.37; 40.91]	[26.65; 42.73]	[26.73; 42.77]
feed=opp	9.44*	9.63*	8.68	8.61
	[2.66; 16.33]	[2.78; 16.64]	[−0.88; 18.09]	[−1.06; 18.23]
prompt=emp	2.65	5.23	1.59	3.88
	[−3.65; 9.07]	[−1.26; 11.70]	[−9.43; 12.62]	[−7.19; 15.03]
feed=opp × *prompt=emp*			1.48	1.96
			[−11.47; 14.47]	[−11.11; 14.84]
days active	−7.24	−6.96	−7.20	−6.83
	[−14.84; 0.22]	[−14.64; 0.59]	[−14.69; 0.31]	[−14.54; 0.88]
statuses count	10.18	9.43	10.02	9.30
	[−3.51; 23.80]	[−4.11; 23.03]	[−3.51; 23.52]	[−4.66; 23.08]
favorites count	−4.07	−3.53	−4.06	−3.44
	[−13.84; 5.67]	[−13.38; 6.31]	[−13.89; 5.82]	[−13.42; 6.52]
followers count	1.49	2.01	1.34	1.74
	[−7.90; 10.85]	[−7.52; 11.75]	[−8.34; 11.07]	[−8.08; 11.63]
friends count	−10.33*	−11.13*	−10.28*	−11.09*
	[−19.98; −0.62]	[−20.91; −1.28]	[−19.94; −0.67]	[−21.09; −1.15]
feed=opp × *days active*	5.41	4.99	5.35	4.87
	[−2.10; 12.91]	[−2.63; 12.64]	[−2.07; 12.84]	[−2.81; 12.57]
feed=opp × *statuses count*	−17.92*	−16.90*	−17.64*	−16.70*
	[−32.31; −3.50]	[−31.15; −2.47]	[−32.09; −3.20]	[−31.21; −1.67]
feed=opp × *favorites count*	0.15	−0.74	0.05	−0.89
	[−10.55; 10.81]	[−11.60; 10.12]	[−10.66; 10.61]	[−11.66; 10.08]
feed=opp × *followers count*	3.41	2.46	3.61	2.86
	[−7.51; 14.44]	[−8.67; 13.44]	[−7.64; 15.00]	[−8.53; 14.22]
feed=opp × *friends count*	9.72	11.12*	9.58	10.92
	[−0.91; 20.38]	[0.17; 21.97]	[−1.14; 20.36]	[−0.02; 21.95]
prompt=emp × *days active*	1.56	1.21	1.58	1.18
	[−5.43; 8.58]	[−5.83; 8.29]	[−5.34; 8.55]	[−5.86; 8.34]
prompt=emp × *statuses count*	−10.63	−8.85	−10.80	−8.73
	[−23.96; 2.68]	[−22.27; 4.36]	[−23.78; 2.51]	[−22.11; 4.68]
prompt=emp × *favorites count*	12.51*	12.54*	12.74*	12.61*
	[2.34; 22.65]	[2.27; 22.62]	[2.55; 22.87]	[2.36; 22.90]
prompt=emp × *followers count*	1.88	1.74	1.85	1.52
	[−9.27; 13.03]	[−9.55; 13.08]	[−9.54; 13.09]	[−10.00; 13.04]
prompt=emp × *friends count*	3.82	4.24	3.91	4.42
	[−6.29; 13.86]	[−6.04; 14.42]	[−6.39; 14.19]	[−6.03; 14.74]
Num. obs.	242.00	242.00	242.00	242.00

* 0 outside the credible interval.

Table S21: BRMS monotonicity estimates for prompt ATE for *Survey Question 2*.

	Treatment: Prompt, Outcome: Survey Q3 (”I can understand why some people might identify with the views shown in this feed”)			
	lower bound	*upper bound*	*lower bound w/ interactions*	*upper bound w/ interactions*
(Intercept)	55.40*	54.94*	55.77*	55.75*
	[48.25; 62.43]	[48.05; 61.83]	[47.20; 64.26]	[47.54; 64.00]
feed=opp	−10.62*	−10.02*	−11.21*	−11.21*
	[−17.85; −3.41]	[−17.22; −2.86]	[−21.19; −1.24]	[−20.83; −1.53]
prompt=emp	7.53*	8.61*	6.75	6.94
	[0.94; 14.12]	[2.23; 15.07]	[−4.95; 18.56]	[−4.45; 18.36]
feed=opp × *prompt=emp*			1.11	2.38
			[−12.32; 14.51]	[−10.67; 15.55]
days active	4.00	3.26	4.02	3.25
	[−3.76; 11.94]	[−4.26; 10.83]	[−3.83; 11.83]	[−4.34; 10.90]
statuses count	−12.17	−11.94	−12.19	−11.83
	[−26.17; 2.19]	[−25.80; 1.96]	[−26.26; 2.10]	[−25.86; 1.94]
favorites count	7.40	6.96	7.38	6.87
	[−3.25; 18.12]	[−3.31; 17.23]	[−3.14; 17.77]	[−3.49; 17.32]
followers count	4.98	5.28	4.91	5.01
	[−4.75; 14.76]	[−4.25; 14.95]	[−4.98; 14.75]	[−4.55; 14.74]
friends count	3.88	4.29	3.93	4.43
	[−6.20; 14.07]	[−5.70; 14.29]	[−6.25; 14.08]	[−5.62; 14.50]
feed=opp × *days active*	0.95	1.98	0.96	1.99
	[−6.85; 8.55]	[−5.50; 9.31]	[−6.69; 8.62]	[−5.54; 9.43]
feed=opp × *statuses count*	−0.66	−0.99	−0.50	−0.95
	[−15.94; 14.57]	[−15.87; 13.86]	[−15.72; 14.84]	[−15.68; 13.92]
feed=opp × *favorites count*	−2.16	−1.42	−2.17	−1.43
	[−13.42; 9.11]	[−12.25; 9.61]	[−13.37; 9.20]	[−12.32; 9.41]
feed=opp × *followers count*	−3.62	−4.02	−3.55	−3.71
	[−14.89; 7.40]	[−14.98; 6.95]	[−14.86; 7.67]	[−14.73; 7.44]
feed=opp × *friends count*	−3.64	−4.24	−3.75	−4.44
	[−14.85; 7.46]	[−14.94; 6.56]	[−14.89; 7.52]	[−15.49; 6.54]
prompt=emp × *days active*	−5.84	−4.64	−5.88	−4.60
	[−12.85; 1.18]	[−11.47; 2.26]	[−12.87; 1.11]	[−11.43; 2.28]
prompt=emp × *statuses count*	5.71	5.23	5.67	5.09
	[−7.54; 18.85]	[−7.67; 17.76]	[−7.46; 18.94]	[−7.95; 18.17]
prompt=emp × *favorites count*	−2.96	−2.30	−2.89	−2.05
	[−13.20; 7.29]	[−12.38; 7.65]	[−13.05; 7.52]	[−12.35; 8.16]
prompt=emp × *followers count*	−0.20	−0.26	−0.27	−0.40
	[−11.20; 10.76]	[−10.93; 10.51]	[−11.24; 10.53]	[−11.37; 10.41]
prompt=emp × *friends count*	−3.42	−4.32	−3.36	−4.26
	[−13.73; 6.80]	[−14.23; 5.78]	[−13.71; 6.96]	[−14.53; 6.05]
Num. obs.	235.00	233.00	235.00	233.00

* 0 outside the credible interval.

Table S22: BRMS monotonicity estimates for prompt ATE for *Survey Question 3*.

	Treatment: Prompt, Outcome: Survey Q4 (”In the future, I would be interested in having a conversation with this feed’s owner”)			
	lower bound	*upper bound*	*lower bound* *w/ interactions*	*upper bound* *w/ interactions*
(Intercept)	28.96*	28.75*	24.39*	25.01*
	[21.07; 36.83]	[20.50; 37.08]	[15.02; 33.85]	[15.66; 34.59]
feed=opp	9.79*	10.06*	16.37*	15.49*
	[1.56; 17.91]	[1.31; 18.88]	[5.24; 27.52]	[4.52; 26.45]
prompt=emp	6.71	11.07*	15.88*	19.86*
	[−0.80; 14.25]	[3.35; 18.63]	[3.05; 28.76]	[6.64; 33.14]
feed=opp × *prompt=emp*			−12.98	−12.27
			[−28.15; 2.02]	[−27.29; 2.96]
days active	−11.46*	−10.82*	−11.79*	−10.60*
	[−20.38; −2.52]	[−20.01; −1.58]	[−20.76; −2.91]	[−19.83; −1.19]
statuses count	3.99	0.50	3.73	−0.88
	[−12.18; 20.03]	[−16.57; 17.74]	[−12.41; 19.80]	[−18.31; 16.23]
favorites count	2.26	2.74	2.16	2.69
	[−9.30; 13.95]	[−8.76; 14.13]	[−9.52; 13.81]	[−8.81; 14.27]
followers count	8.22	8.79	9.70	10.24
	[−5.71; 22.04]	[−5.27; 22.63]	[−4.48; 23.49]	[−4.04; 24.44]
friends count	−3.24	−3.81	−3.46	−4.08
	[−15.49; 9.21]	[−16.28; 8.53]	[−15.92; 8.91]	[−16.48; 8.27]
feed=opp × *days active*	1.93	1.43	1.94	0.85
	[−6.93; 10.63]	[−7.68; 10.62]	[−6.82; 10.70]	[−8.42; 10.08]
feed=opp × *statuses count*	0.08	4.48	−0.36	5.35
	[−16.57; 16.61]	[−14.12; 22.73]	[−17.08; 16.42]	[−12.85; 23.82]
feed=opp × *favorites count*	−0.32	−0.93	−0.20	−0.70
	[−12.60; 12.13]	[−13.12; 11.28]	[−12.74; 11.99]	[−13.11; 11.68]
feed=opp × *followers count*	−15.84*	−16.59*	−17.35*	−18.03*
	[−30.24; −1.39]	[−31.19; −1.87]	[−31.85; −2.73]	[−33.05; −2.89]
feed=opp × *friends count*	5.68	6.50	6.36	7.08
	[−7.42; 18.90]	[−6.74; 19.84]	[−6.89; 19.50]	[−6.10; 20.32]
prompt=emp × *days active*	0.25	2.10	0.50	2.46
	[−7.82; 8.23]	[−5.97; 10.15]	[−7.61; 8.65]	[−5.61; 10.38]
prompt=emp × *statuses count*	−13.09	−11.68	−12.63	−11.89
	[−27.83; 1.92]	[−26.91; 3.57]	[−27.35; 2.14]	[−26.90; 3.13]
prompt=emp × *favorites count*	12.74*	10.09	12.39*	9.65
	[1.13; 24.33]	[−1.44; 21.56]	[0.80; 24.05]	[−1.72; 21.12]
prompt=emp × *followers count*	6.28	4.53	6.73	4.98
	[−6.59; 18.79]	[−8.38; 17.58]	[−6.29; 19.75]	[−8.00; 17.89]
prompt=emp × *friends count*	−4.14	−1.78	−4.60	−2.15
	[−15.87; 7.49]	[−13.82; 10.12]	[−16.00; 6.88]	[−14.10; 9.73]
Num. obs.	235.00	222.00	235.00	222.00

* 0 outside the credible interval.

Table S23: BRMS monotonicity estimates for prompt ATE for *Survey Question 4*.

References

[1] Bates, D., Mächler, M., Bolker, B. M. & Walker, S. C. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* **67** (2015).

[2] Bürkner, P.-C. brms: An r package for bayesian multilevel models using stan. *Journal of Statistical Software, Articles* **80**, 1–28 (2017).

[3] Bürkner, P.-C. Advanced bayesian multilevel modeling with the r package brms. *R Journal* **10** (2018).

[4] Barr, D. J., Levy, R., Scheepers, C. & Tily, H. J. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language* **68**, 255–278 (2013).

[5] Rosenbaum, P. R. *Design of observational studies*, vol. 10 (Springer, 2010).

[6] Gerber, A. S. & Green, D. *Field Experiments: Design, Analysis, and Interpretation* (New York: W. W. Norton & Company, 2012).