Supplemental Materials

Evaluations of Empathizers Depend on the Target of Empathy

Y. Andre Wang

Andrew R. Todd

University of California, Davis

Journal of Personality and Social Psychology

https://dx.doi.org/10.1037/pspi0000341.supp

Table of Contents

Experiment 1

Participants and Power Analysis

Amazon's Mechanical Turk (MTurk) workers ($N = 464$) completed the experiment online for modest remuneration. Because this was the first experiment in this line of research, we did not have an effect size estimate and thus set a target sample size that would provide 80% power ($\alpha = .05$) to detect a small effect ($\eta_p^2 = .02$) in a 2 × 2 between-subjects design. A power analysis[1] suggested a minimum sample size of $N = 387$. We anticipated an exclusion rate of 10% and decided to collect data from 450 participants.[2] We decided a priori on the following exclusion criteria: failing the attention check on Beth's response to Ann, failing the attention check on Ann's employer, and giving identical responses across all dependent variables (DVs; because we had reverse-coded items). Upon concluding data collection but prior to analysis, we decided to retain data from participants who gave identical neutral responses (i.e., 4 on a 7-point scale) across the DVs; we reasoned that one could plausibly feel neutral on all items (this modification did not change our results). The numbers of participants who met each exclusion criterion were 89, 52, and 7, respectively. The final sample size was 336 (some participants met more than one criterion).[3]

Manipulation Checks

The manipulations of target valence and response type were both successful: Participants in the positive target (vs. negative target) conditions viewed Ann more positively ($M = 5.83$, SD

[1] Across experiments, we conducted all power analyses with G*Power (Faul, Erdfelder, Lang, & Buchner, 2007), unless otherwise noted.
[2] The number of participants initially included in this and subsequent samples was slightly higher than our target sample size because our data collection platform counted the number of people who proceeded to the last page of our experiment rather than the number of people who completed all survey questions.
[3] Due to a programming oversight, we did not collect information on participant gender and age in Experiments 1 and S1. We report participant gender and age for all other experiments.

= 1.00 vs. M = 1.97, SD = 1.58), $t(255) = 26.37$, $p < .001$, $d = 2.97$, 95% confidence interval (CI) [2.66, 3.29], and participants in the empathic (vs. nonempathic) response conditions thought that Beth empathized with Ann to a greater extent (M = 5.83, SD = 1.15 vs. M = 4.03, SD = 1.46), $t(263) = 12.26$, $p < .001$, $d = 1.40$, 95% CI [1.16, 1.64].

Experiment S1

Experiment S1 was a conceptual replication of Experiment 1. We used the same experimental design but extended the dialogue to clarify that Beth and Ann did not know each other beforehand. The dialogue also included Beth's confirmation that she knew where Ann worked, thus removing ambiguity about whether Beth understood the mission of Ann's employer. Despite these changes, we expected to replicate the results of Experiment 1.

Method

Participants and power analysis. MTurk workers (N = 472) participated online for modest remuneration. A power analysis indicated that a sample size of N = 296 affords 80% power ($\alpha = .05$) to detect an effect comparable in size to the key interaction effects in Experiment 1 (around $\eta_p^2 = .026$). Given the 28% exclusion rate in Experiment 1 and the need to exclude participants who had completed Experiment 1, we decided to match the sample size of Experiment 1 and collect data from 450 participants. As in Experiment 1, we decided a priori to exclude data from participants who failed the attention check on Beth's response to Ann, failed the attention check on Ann's employer, or gave identical nonneutral responses (i.e., other than 4 on 7-point scales) across all DVs. We also decided a priori to exclude data from participants who indicated that they had completed Experiment 1.[4] The numbers of participants who met each

[4] In subsequent experiments, only MTurk workers who had not already participated in a study in this line of research were eligible to participate, so this data exclusion criterion was not used in Experiments 2 through 7.

criterion were 56, 43, 5, and 48, respectively. The final sample size was 373 (some participants met more than one exclusion criterion).

Materials and procedure. This experiment was identical to Experiment 1, except for the differences reported in the following text. Participants (1) learned that Beth and Ann were meeting for the first time at a neighborhood dog park and (2) read a more extensive dialogue, during which Ann revealed the organization she worked for (text for the positive target conditions appears below; in the negative target conditions, the organization name was replaced with "Aryan Nations"):

> Beth: "I don't think I've ever seen you around here before. Are you new to the neighborhood?"
>
> Ann: "Yes, I just moved here. My name is Ann. Nice to meet you."
>
> Beth: "Nice to meet you! I'm Beth. How are you doing?"
>
> Ann: "Well...not so great, to be honest."
>
> Beth: "How come?"
>
> Ann: "I'm feeling really stressed. I work for this organization called St. Jude Children's Research Hospital. Are you familiar with it?"
>
> Beth: "Yes, I've heard of it."
>
> Ann: "So yeah, I do event planning and outreach for St. Jude Children's Research Hospital, and I'm organizing an event for them. My team is expecting a large attendance, but I've been having a lot of trouble with the logistics of it, and the date of the event was recently delayed because we did not hear back from the city council in time. The stress is overwhelming and has affected my sleep, and I've been feeling awful because of it."

The options for the attention check on Beth's response to Ann included Beth's full responses ("I feel for you—I can really put myself in your shoes in this situation. When is the event taking place?", "Okay, I see. When is the event taking place?", "I don't understand your situation. When is the event taking place?", or "none of the above").

Results

Manipulation checks. Both manipulations were again successful: Participants in the positive (vs. negative) target conditions viewed Ann more positively (M = 5.82, SD = 1.08 vs. M = 1.82, SD = 1.46), $t(346) = 30.17$, $p < .001$, $d = 3.11$, 95% CI [2.81, 3.42], and participants in the empathic (vs. nonempathic) response conditions indicated that Beth empathized with Ann to a greater extent (M = 5.77, SD = 1.25 vs. M = 4.35, SD = 1.42), $t(365) = 10.23$, $p < .001$, d = 1.06, 95% CI [0.84, 1.28].

Factor analysis. To confirm the factor structure from Experiment 1, we conducted a confirmatory factor analysis (CFA) in R using the *lavaan* package (Rosseel, 2012). Drawing from the EFA solution in Experiment 1, we specified a model with two latent factors; four items (like, respect, trust, and friends) loaded onto the first factor (respect/liking), and the other four items (understanding, kind, cold [reverse-coded], and caring) loaded onto the second factor (warmth). Because factor loadings of all items on their nonprimary factors were low in the EFA solution in Experiment 1, we specified no cross-loadings in the CFA. This two-factor model fit the data well, $\chi^2(19) = 59.96$, $p < .001$, root mean square error of approximation (RMSEA) = 0.08, confirmatory fit index (CFI) = 0.99, Tucker–Lewis index (TLI) = 0.98, with all factor loadings higher than $\lambda = .60$. The two-factor model also fit the data better than a one-factor model in which all items loaded onto a single factor, $\Delta\chi^2(1) = 184.57$, $p < .001$. Thus, we confirmed the factor structure from Experiment 1. As in Experiment 1, we calculated the mean ratings of items for respect/liking ($\alpha = .96$) and warmth ($\alpha = .90$) as composites and conducted the primary analyses on those composites.

Respect/liking. A 2 (response type) × 2 (target valence) between-subjects ANOVA on respect/liking again yielded main effects of both factors: Participants respected/liked Beth more

when she gave an empathic (vs. nonempathic) response, $F(1, 369) = 7.44$, $p = .007$, $\eta_p^2 = .02$, 90% CI [.003, .05], and when Ann was positively (vs. negatively) portrayed, $F(1, 369) = 123.73$, $p < .001$, $\eta_p^2 = .25$, 90% CI [.19, .31]. The Response Type × Target Valence interaction was marginally significant, $F(1, 369) = 3.13$, $p = .078$, $\eta_p^2 = .01$, 90% CI [.00, .03]. When Ann was positively portrayed, participants respected/liked Beth more when she gave an empathic (vs. nonempathic) response ($M = 5.57$, $SD = 1.13$ vs. $M = 4.94$, $SD = 0.96$), $F(1, 369) = 9.96$, $p = .002$, $\eta_p^2 = .03$, 90% CI [.01, .06]. When Ann was negatively portrayed, however, respect/liking for Beth did not significantly differ by response type ($M = 3.77$, $SD = 1.72$ vs. $M = 3.64$, $SD = 1.44$), $F(1, 369) = 0.47$, $p = .495$, $\eta_p^2 < .01$, 90% CI [.00, .01] (see Figure S1, left panel).

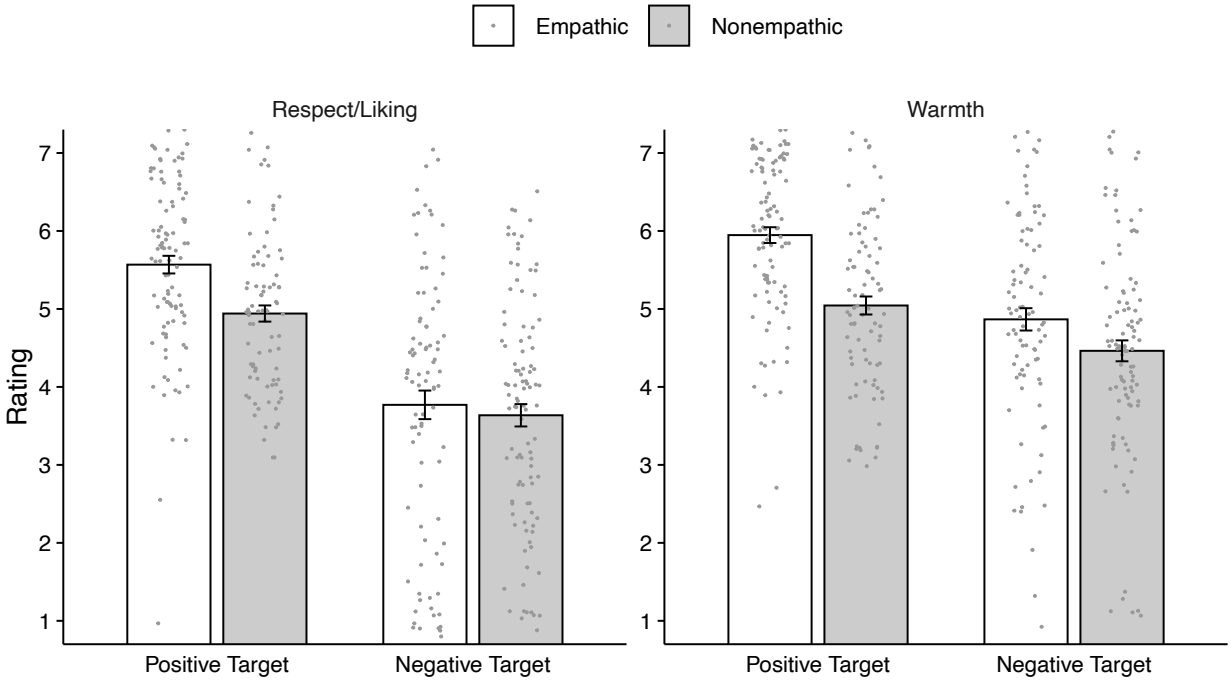

Figure S1. Ratings of Beth on respect/liking and warmth by response type and target valence in Experiment S1. Error bars depict ±1 standard errors; dots depict jittered individual data points.

Warmth. An identical 2 × 2 ANOVA on warmth also revealed main effects of response type and target valence: Participants rated Beth as warmer when she gave an empathic (vs. nonempathic) response, $F(1, 369) = 27.39$, $p < .001$, $\eta_p^2 = .07$, 90% CI [.03, .11], and when Ann

was positively (vs. negatively) portrayed, $F(1, 369) = 44.32$, $p < .001$, $\eta_p^2 = .11$, 90% CI [.06, .16]. The Response Type × Target Valence interaction was (barely) significant, $F(1, 369) = 4.00$, $p = .046$, $\eta_p^2 = .01$, 90% CI [.0001, .03]. When Ann was positively portrayed, participants rated Beth as warmer when she gave an empathic (vs. nonempathic) response ($M = 5.95$, $SD = 1.01$ vs. $M = 5.04$, $SD = 1.06$), $F(1, 369) = 25.80$, $p < .001$, $\eta_p^2 = .07$, 90% CI [.03, .11]. Unlike the respect/liking results, participants rated Beth as warmer when she gave an empathic (vs. nonempathic) response to negatively portrayed Ann ($M = 4.87$, $SD = 1.34$ vs. $M = 4.46$, $SD = 1.35$), though the effect was smaller, $F(1, 369) = 5.30$, $p = .022$, $\eta_p^2 = .01$, 90% CI [.001, .04] (see Figure S1, right panel).

Discussion

Experiment S1 largely replicated the results of Experiment 1. First, we confirmed the same two-dimensional structure of evaluations of the responder. Second, using an extended, less ambiguous dialogue, we again found that evaluations of empathizers depended on the empathic target (though the Response Type × Target Valence interaction on respect/liking was marginally significant). Participants respected/liked the responder more when she responded empathically to a positively portrayed target, but not when she responded to a negatively portrayed target. Participants also rated the responder as warmer when she responded empathically, but this effect was smaller when the target was negatively portrayed. Although the effect sizes were smaller here than in Experiment 1, the overall pattern of results was largely unaffected by assumptions about Beth and Ann's relationship or Beth's knowledge about Ann's employer.

Experiment 2

Participants and Power Analysis

We publicly pre-registered our analysis plan, including power analyses, target sample size, inclusion and exclusion criteria, and planned data analyses, on AsPredicted (https://aspredicted.org/qr8nk.pdf). MTurk workers (N = 614, 49% female, 51% male; M_{age} = 37.9, SD = 12.5) participated online for modest remuneration. We powered this experiment at 80% (α = .05) to detect the expected effect of Target Valence × Response Type interaction on respect/liking and warmth. Based on past experiments, we estimated the interaction effect sizes as η_p^2 = .017 (respect/liking) and η_p^2 = .029 (warmth). The sample sizes required to detect these effect sizes are 456 and 265. We chose the more conservative N = 456 as the target sample size for analysis. Based on an anticipated exclusion rate of 25% estimated from previous experiments, we set a target sample size of 608. We decided a priori to exclude participants based on the same three exclusion criteria from Experiment 1. The numbers of participants who met each exclusion criterion were 33, 70, and 2, respectively. The final sample size was 526 (some participants met more than one exclusion criterion).

Manipulation Checks

All manipulations were successful. Participants evaluated Ann more positively when she was portrayed as pro- versus anti-vaccination (M = 5.66, SD = 1.19 vs. M = 2.80, SD = 1.74), $t(442) = 21.94$, $p < .001$, d = 1.94, 95% CI [1.73, 2.15]. As expected, this target valence manipulation was considerably weaker than the manipulation used in Experiments 1 and 2 (recall that the effect sizes on the manipulation check were d = 2.97 and d = 3.11, respectively). Participants also indicated that Beth empathized with Ann to a greater extent when she gave an empathic (vs. nonempathic) response (M = 5.61, SD = 1.29 vs. M = 2.95, SD = 1.55), $t(506) = 21.45$, $p < .001$, d = 1.87, 95% CI [1.67, 2.08], and they rated Ann as feeling negative at the

beginning of the interaction (M = 2.92, SD = 1.69), with the mean significantly below the mid-point of the scale, $t(525) = 14.67$, $p < .001$, $d = 0.64$, 95% CI [0.55, 0.73].

Experiment 3

Participants and Power Analysis

MTurk workers (N = 507, 52% female, 44% male, 4% no gender information; M_{age} = 37.4, SD_{age} = 12.6) participated online for modest remuneration. We determined our target sample size by running two power analyses based on the effect size estimates of the key interactions in Experiments 1 (around $\eta_p^2 = .026$) and 2 (around $\eta_p^2 = .011$). Detecting these two effect sizes at 80% power ($\alpha = .05$) would require sample sizes of 296 and 708, respectively. Because we were unsure which effect size was more likely, and because we anticipated an exclusion rate of around 19% (based on the average exclusion rate in Experiments 1 and 2), we set a target sample size toward the higher end of the sample sizes suggested by the power analyses and collected data from 500 participants. As in Experiments 1 and 2, we decided a priori on the following exclusion criteria: failing the attention check on Beth's response to Ann, failing the attention check on Ann's employer, and giving identical non-neutral responses (i.e., other than 4 on 7-point scales) across all DVs. The numbers of participants who met each criterion were 33, 65, and 6, respectively. The final sample size was 416 (some participants met more than one criterion).

Manipulation Checks

All manipulations were successful: Participants evaluated Ann more positively when she was positively (vs. negatively) portrayed (M = 5.87, SD = 1.05 vs. M = 2.01, SD = 1.49), $t(323) = 29.87$, $p < .001$, $d = 3.05$, 95% CI [2.76, 3.33]. They also indicated that Beth empathized with Ann to a greater extent when she gave an empathic (vs. nonempathic) response (M = 5.59, SD =

1.21 vs. $M = 2.67$, $SD = 1.39$), $t(404) = 22.80$, $p < .001$, $d = 2.24$, 95% CI [1.99, 2.48]. In addition, participants rated Ann as feeling positive at the beginning of the interaction ($M = 6.20$, $SD = 1.22$), with the mean significantly above the mid-point of the scale, $t(415) = 36.67$, $p < .001$, $d = 1.80$, 95% CI [1.64, 1.95].

Experiment 4

Participants and Power Analysis

We publicly pre-registered our analysis plan, including power analyses, target sample size, inclusion and exclusion criteria, and planned data analyses, on AsPredicted (https://aspredicted.org/uq3ct.pdf). MTurk workers ($N = 838$, 58% female, 42% male; $M_{age} = 39.6$, $SD = 12.4$) participated online for modest remuneration. We powered this experiment at 80% ($\alpha = .05$) to detect the expected effect size of response type on the primary DVs in the positive target condition. A conservative estimate of $\eta_p^2 = .031$ from simple effects analyses in the previous experiments suggested a target sample size of 124 per condition ($N = 744$). Based on an anticipated exclusion rate of 10% estimated from previous experiments, we set a target sample size of 820. We decided a priori to exclude participants based on the same three exclusion criteria from Experiment 3.[5] The numbers of participants who met each exclusion criterion were 45, 63, and 1, respectively. The final sample size was 740 (some participants met more than one exclusion criterion).

Manipulation Checks

Following our pre-analysis plan, we conducted a one-tailed independent samples *t*-test on the manipulation check of target valence. Participants viewed Ann more positively when she was

[5] We had reported an additional exclusion criterion in the pre-registration: excluding participants who fail the captcha verification at the beginning of the experiment. In reality, because the captcha verification appeared before any data could be recorded, all participants with recorded data passed the captcha verification.

positively (vs. negatively) portrayed (M = 5.77, SD = 1.06 vs. M = 1.94, SD = 1.36), $t(680)$ = 42.53, $p < .001$, d = 3.15, 95% CI [2.93, 3.36]. In addition, a one-tailed one sample t-test on Ann's affect confirmed that participants rated Ann as feeling negative at the beginning of the interaction (M = 2.35, SD = 1.37), with the mean significantly below the midpoint of the scale, $t(740) = 32.81$, $p < .001$, d = 1.21, 95% CI [1.11, 1.30]. Furthermore, an exploratory analysis indicated that, as in the pilot study, participants in the positive empathic response condition thought Beth empathized with Ann significantly more than did participants in both the positive nonempathic response condition (M = 5.52, SD = 1.33 vs. M = 4.19, SD = 1.79), $t(463) = 9.44$, $p < .001$, d = 0.84, 95% CI [0.66, 1.02], and the neutral nonempathic condition (M = 2.82, SD = 1.43), $t(480) = 21.51$, $p < .001$, d = 1.95, 95% CI [1.73, 2.16].

Fit of Latent Mediation Models

A summary of model fit indices is reported in Table S1.

Table S1

Summary of Latent Moderated Mediation Models Tested in Experiment 4.

Model	Planned or Exploratory?	Predictor	DV	χ^2	CFI	TFI	RMSEA
1	Planned	Empathy	Respect/liking	269.86	0.97	0.96	0.07
2	Planned		Warmth	358.81	0.95	0.94	0.08
3	Planned	Positivity	Respect/liking	513.89	0.94	0.92	0.10
4	Planned		Warmth	580.33	0.93	0.90	0.11
5	Exploratory	Empathic vs. positive nonempathic	Respect/liking	272.33	0.95	0.94	0.09
6	Exploratory		Warmth	383.28	0.91	0.89	0.11

Note. Although the fit indices of some models slightly differed from conventional recommendations, inspection of residual matrices suggested that all models fit the data reasonably well and that the fit indices were oversensitive to minor model misspecifications, given the low unique variances of some observed variables (< .10; Browne, MacCallum, Kim, Andersen, & Glaser, 2002). In all models, df = 57, ps < .001.

Descriptions of Moderated Mediation Models 3 Through 6

Moderated mediation models with response positivity as predictor (Models 3 and 4). In Models 3 and 4, we conducted our planned moderated mediation analyses using response positivity as the predictor. Analysis on respect/liking (Model 3) indicated that response positivity significantly predicted the mediator (a = 1.78, $p < .001$), and that the Mediator × Target Valence interaction significantly predicted respect/liking (b_{mod} = 0.54, $p < .001$), suggesting the presence of second-stage moderated mediation. However, because the Response Positivity × Target Valence interaction did not significantly predict the mediator (a_{mod} = -0.08, p = .355), there was no evidence of first-stage moderated mediation. Supporting these results, the effect of response positivity on inferences about Beth's attitudes towards Ann was similar across target valence (a_{pos} = 1.70 vs. a_{neg} = 1.86), but the association between the mediator and respect/liking was stronger when Ann was positively portrayed (b_{pos} = 0.97 vs. b_{neg} = -0.10), and the overall indirect effect was also stronger when Ann was positively portrayed ($a_{pos}b_{pos}$ = 1.65 vs. $a_{neg}b_{neg}$ = -0.18).

We then conducted the same analysis on warmth (Model 4), and the results were similar. We again saw evidence of second-stage moderated mediation, in which response positivity significantly predicted the mediator (a), and the Mediator × Target Valence interaction significantly predicted warmth (b_{mod} = 0.58, $p < .001$). Because a_{mod} was not significant, there was again no evidence of first-stage moderated mediation. Supporting these results, the effect of response positivity on inferences about Beth's attitudes toward Ann was similar across target valence (a_{pos} = 1.68 vs. a_{neg} = 1.85), but the association between the mediator and warmth was stronger when Ann was positively portrayed (b_{pos} = 1.23 vs. b_{neg} = 0.07), and the overall indirect effect was also stronger when Ann was positively portrayed ($a_{pos}b_{pos}$ = 2.06 vs. $a_{neg}b_{neg}$ = 0.14).

Taken together, Models 3 and 4 indicated that second-stage moderated mediation was present when we compared the effects of positive versus neutral responses: Inferences about

Beth's attitudes toward Ann mediated the Response Positivity × Parget Valence interaction on evaluations of Beth, but such inferences were predicted only by response positivity and did not differ by target valence.

Moderated mediation models with empathic vs. positive nonempathic response as predictor (Models 5 and 6). We explored within the empathic and positive nonempathic response conditions whether the Response Type × Target Valence interaction on evaluations of Beth was mediated. Analysis on respect/liking showed that the Response Type × Target Valence interaction significantly predicted the mediator (a_{mod} = 0.25, p < .001), and that the mediator significantly predicted respect/liking (b = 0.37, p < .001), suggesting the presence of first-stage moderated mediation (Model 5). In addition, response type significantly predicted the mediator (a = 0.18, p = .001), and the Mediator × Target Valence interaction significantly predicted respect/liking (b_{mod} = 0.41, p < .001), suggesting the presence of second-stage moderated mediation as well. Supporting these results, the effect of response type on inferences about Beth's attitude toward Ann, the association between those inferences and respect/liking, and the overall indirect effect were all stronger when Ann was positively portrayed (a_{pos} = 0.43 vs. a_{neg} = -0.06, b_{pos} = 0.78 vs. b_{neg} = -0.04, $a_{pos}b_{pos}$ = 0.34 vs. $a_{neg}b_{neg}$ = 0.00).

The same exploratory analysis on warmth showed highly similar results: Response Type × Target Valence interaction significantly predicted the mediator (a_{mod}), and the mediator significantly predicted warmth (b = 0.57, p < .001), suggesting the presence of first-stage moderated mediation (Model 6). In addition, response type significantly predicted the mediator (a), and the Mediator × Target Valence interaction significantly predicted warmth (b_{mod} = 0.42, p < .001), suggesting the presence of second-stage moderated mediation as well. Supporting these results, the effect of response type on inferences about Beth's attitude toward Ann, the

associations between those inferences and warmth, and the overall indirect effects were all stronger when Ann was positively portrayed (a_{pos} = 0.43 vs. a_{neg} = -0.06, b_{pos} = 0.99 vs. b_{neg} = 0.14, $a_{pos}b_{pos}$ = 0.43 vs. $a_{neg}b_{neg}$ = -0.01). Taken together, Models 5 and 6 suggested that evidence of both first and second-stage moderated mediation was present even when comparing only the effects of empathic versus positive nonempathic responses: Inferences about Beth's attitude toward Ann mediated the Response Type × Target Valence interaction on evaluations of Beth, and those inferences were predicted by the Response Type × Target Valence interaction.

Experiment 5

Participants and Power Analysis

MTurk workers (N = 504, 50% female, 41% male, 9% no gender information; M_{age} = 39.0, SD_{age} = 11.9) participated online for modest remuneration. We powered our experiment to detect two effects: the effect of response type on respect/liking, and the indirect effect of the mediator on the DVs. Our experimental design was similar to that of the negative target conditions in Experiment 1; however, we reasoned that the condemning (vs. empathic) response should have a larger effect than the nonempathic (vs. empathic) response. Therefore, we estimated the effect size of response type on respect/liking as d = 0.28, which was twice as large as the size of the simple effect of nonempathic (vs. empathic) response on respect/liking in Experiment 1 (d = 0.14). Powering this experiment to detect an effect size of d = 0.28 at 80% (α = .05) requires N = 404. This sample size also affords > 80% power to detect an indirect effect as small as ab = 0.03, based on simulations using the power analysis app for mediation models developed by Schoemann, Boulton, and Short (2017).[6] Using a conservative estimate of 20%

[6] We conducted a power analysis using observed mediation models instead of our planned latent mediation models due to challenges of conducting power analysis for the latter. Because we anticipated that our latent variables would be highly reliable (αs = .90–.95), however, using latent variables in our mediation models should result in negligible power loss (see Table 3 in Ledgerwood & Shrout, 2011; see also Wang & Rhemtulla, in press).

exclusion rate, we aimed to collect data from 500 participants. We decided a priori on the same three exclusion criteria used in Experiments 3 and 4. The numbers of participants who met each criterion were 22, 37, and 0, respectively. The final sample size was 452 (some participants met more than one exclusion criterion).

Pilot Study

In order to ensure that our response type manipulation was successful, we conducted a pilot study on four candidate responses. In this pilot study (N = 201; 51% female, 39% male, 10% no gender information; M_{age} = 39.5, SD_{age} = 12.2), participants read the same instructions and Ann's experience as those in Experiment 4, but they did not learn any information about Ann. After reading what Ann said, participants then saw four responses from Beth presented in randomized order and rated how positive and how empathic each response was. The four responses were the three responses used in Experiment 5, as well as a neutral empathic response ("Okay, I can understand why you would feel stressed in this situation").

The positive empathic and positive nonempathic responses were comparably positive (M = 5.23, SD = 1.23 vs. M = 5.47, SD = 1.46), $t(200) = 1.94$, $p = .053$, $d = 0.14$, 95% CI [-0.06, 0.33], and more positive than the neutral nonempathic response (M = 3.14, SD = 1.19), ts > 17.85, ps < .001, ds > 1.25. The positive empathic response was also more empathic than both the positive nonempathic response (M = 5.89, SD = 1.23 vs. M = 3.90, SD = 1.91), $t(200) = 12.38$, $p < .001$, $d = 0.87$, 95% CI [0.67, 1.08], and the neutral nonempathic response (M = 2.46, SD = 1.38), $t(199) = 27.91$, $p < .001$, $d = 1.97$, 95% CI [1.73, 2.21]. Because the neutral empathic response was rated almost as positive (M = 4.74, SD = 1.08) as the two positive responses and significantly more positive than the neutral nonempathic response, $t(200) = 15.24$, $p < .001$, d = 1.07, 95% CI [0.87, 1.28], we did not use the neutral empathic response in the main experiment.

Manipulation Checks

All manipulations were successful: Participants viewed Ann negatively (M = 2.02, SD = 1.64), with the mean significantly below the scale mid-point, $t(451) = 25.59$, $p < .001$, $d = 1.20$, 95% CI [1.08, 1.32]. Participants in the empathic (vs. condemning) response conditions indicated that Beth empathized with Ann more (M = 5.62, SD = 1.26 vs. M = 1.51, SD = 1.01), $t(423) = 38.11$, $p < .001$, $d = 3.60$, 95% CI [3.30, 3.90]. Participants also rated Ann as feeling negative at the beginning of the interaction (M = 2.50, SD = 1.55), with the mean significantly below the scale mid-point, $t(451) = 20.53$, $p < .001$, $d = 0.97$, 95% CI [0.85, 1.08].

Experiment 6

Participants and Power Analysis

We publicly pre-registered our analysis plan, including power analyses, target sample size, inclusion and exclusion criteria, and planned data analyses, on AsPredicted (https://aspredicted.org/ud6hh.pdf). MTurk workers (N = 566, 48% female, 52% male; M_{age} = 36.3, SD_{age} = 10.9) participated online for modest remuneration. We powered our experiment to detect a potential effect of Response Type × Character Gender on respect/liking. We estimated the main effect of response type on respect/liking to be $d = 0.50$ (a more conservative estimate than $d = 0.72$ as observed in Experiment 5), which required 64 per cell for 80% power ($\alpha = .05$). The sample size per cell needed to detect a 2 × 2 between-subjects interaction that eliminates the main effect (a "knockout" interaction) is twice the sample size per cell needed to detect the main effect (Giner-Sorolla, 2018; Ledgerwood, 2019), suggesting a target sample size of 512. Based on an anticipated exclusion rate of 10% estimated from previous experiments, we aimed to collect data from 570 participants. We decided a priori on the same three exclusion criteria used in Experiments 3 through 5. The numbers of participants who met each criterion were 86, 115,

and 3, respectively. The final sample size was 404 (some participants met more than one exclusion criterion).

Manipulation Checks

All our manipulations were successful: Participants viewed the target negatively (M = 2.44, SD = 1.94), with the mean significantly below the scale mid-point, $t(403) = 16.25$, $p < .001$, $d = 0.81$, 95% CI [0.70, 0.92]. Participants in the empathic (vs. condemning) response conditions indicated that the responder empathized with the target more ($M = 5.60$, $SD = 1.29$ vs. $M = 1.99$, $SD = 1.58$), $t(387) = 25.17$, $p < .001$, $d = 2.50$, 95% CI [2.24, 2.77]. Participants also rated the target as feeling negative at the beginning of the interaction ($M = 2.84$, $SD = 1.82$), with the mean significantly below the scale mid-point, $t(403) = 12.85$, $p < .001$, $d = 0.64$, 95% CI [0.53, 0.75]. We explored whether character gender inadvertently affected any of the effects above; it did not, ps > .353.

Experiment 7

Participants and Power Analysis

We publicly pre-registered our analysis plan, including power analyses, target sample size, inclusion and exclusion criteria, and planned data analyses, on AsPredicted (https://aspredicted.org/4wj66.pdf). MTurk workers (N = 573, 52% female, 48% male; M_{age} = 36.9, SD_{age} = 11.8) participated online for modest remuneration. We powered this experiment at 80% ($\alpha = .05$) to detect the expected Response Type × Disclosed Experience interaction on respect/liking. Similar to Experiment 6, we estimated the main effect of response type on respect/liking to be $d = 0.50$ (a more conservative estimate than $d = 0.72$ as observed in Experiment 5), which required 64 per cell for 80% power ($\alpha = .05$). The sample size per cell needed to detect a 2 × 2 between-subjects interaction that eliminates the main effect (a

"knockout" interaction) is twice the sample size per cell needed to detect the main effect (Giner-Sorolla, 2018; Ledgerwood, 2019), suggesting a target sample size of 512. Based on an anticipated exclusion rate of 10% estimated from previous experiments, we aimed to collect data from 570 participants. We decided a priori on the same three exclusion criteria used in Experiments 3 through 6. The numbers of participants who met each exclusion criterion were 42, 84, and 0, respectively. The final sample size was 468 (some participants met more than one criterion).

Manipulation Checks

All manipulations were successful: Participants viewed Ann negatively ($M = 2.08$, $SD = 1.61$), with the mean significantly below the mid-point of the scale, $t(467) = 25.77$, $p < .001$, $d = 1.19$, 95% CI [1.07, 1.31]. Participants in the empathic (vs. condemning) response conditions indicated that Beth empathized with Ann to a greater extent ($M = 5.63$, $SD = 1.28$ vs. $M = 1.53$, $SD = 1.18$), $t(461) = 35.91$, $p < .001$, $d = 3.32$, 95% CI [3.04, 3.60]. Participants also rated Ann as feeling negative at the beginning of the interaction ($M = 2.46$, $SD = 1.56$), with the mean significantly below the mid-point of the scale, $t(467) = 21.37$, $p < .001$, $d = 0.99$, 95% CI [0.88, 1.10]. Furthermore, an unplanned, exploratory analysis indicated that participants in the job stress (vs. cancer stress) condition thought Ann's experience was more attributable to the nature of her job ($M = 5.49$, $SD = 1.41$ vs. $M = 2.90$, $SD = 2.09$), $t(408) = 15.73$, $p < .001$, $d = 1.45$, 95% CI [1.25, 1.66]. In other words, we successfully manipulated how strongly the disclosed experience was linked to the source of target valence.

Confirmatory Factor Analysis Results in Experiments 3 Through 7

In each of Experiments 3 through 7, we confirmed the two-factor structure of our DVs by conducting confirmatory factor analyses (CFAs). Diagrams of the models are shown in Figure

S2, and information on model fit is reported in Table S2. In each experiment, we compared the two-factor model to a one-factor model in which all items loaded onto a single factor and found that the two-factor model provided superior fit in each experiment (see Table S2).

Table S2

Summary of CFA Models Tested in Experiments 3–7.

Experiment	Model	χ^2	CFI	TFI	RMSEA	χ^2_{diff}
Experiment 3	Two-factor	103.31	0.98	0.97	0.10	
	One-factor	331.49	0.91	0.88	0.19	228.19
Experiment 4	Two-factor	151.26	0.98	0.97	0.10	
	One-factor	662.24	0.91	0.87	0.21	510.98
Experiment 5	Two-factor	115.70	0.98	0.96	0.11	
	One-factor	639.94	0.84	0.78	0.26	524.24
Experiment 6	Two-factor	115.87	0.97	0.96	0.11	
	One-factor	617.87	0.84	0.77	0.27	509.81
Experiment 7	Two-factor	163.78	0.97	0.95	0.13	
	One-factor	741.00	0.84	0.78	0.28	577.22

Note: In all two-factor models, $df = 19$, $ps < .001$; in all one-factor models, $df = 20$, $ps < .001$. For each experiment, χ^2_{diff} is the chi-square difference between the two-factor model and the one-factor model. In all chi-square difference tests, $df_{diff} = 1$, $ps < .001$.

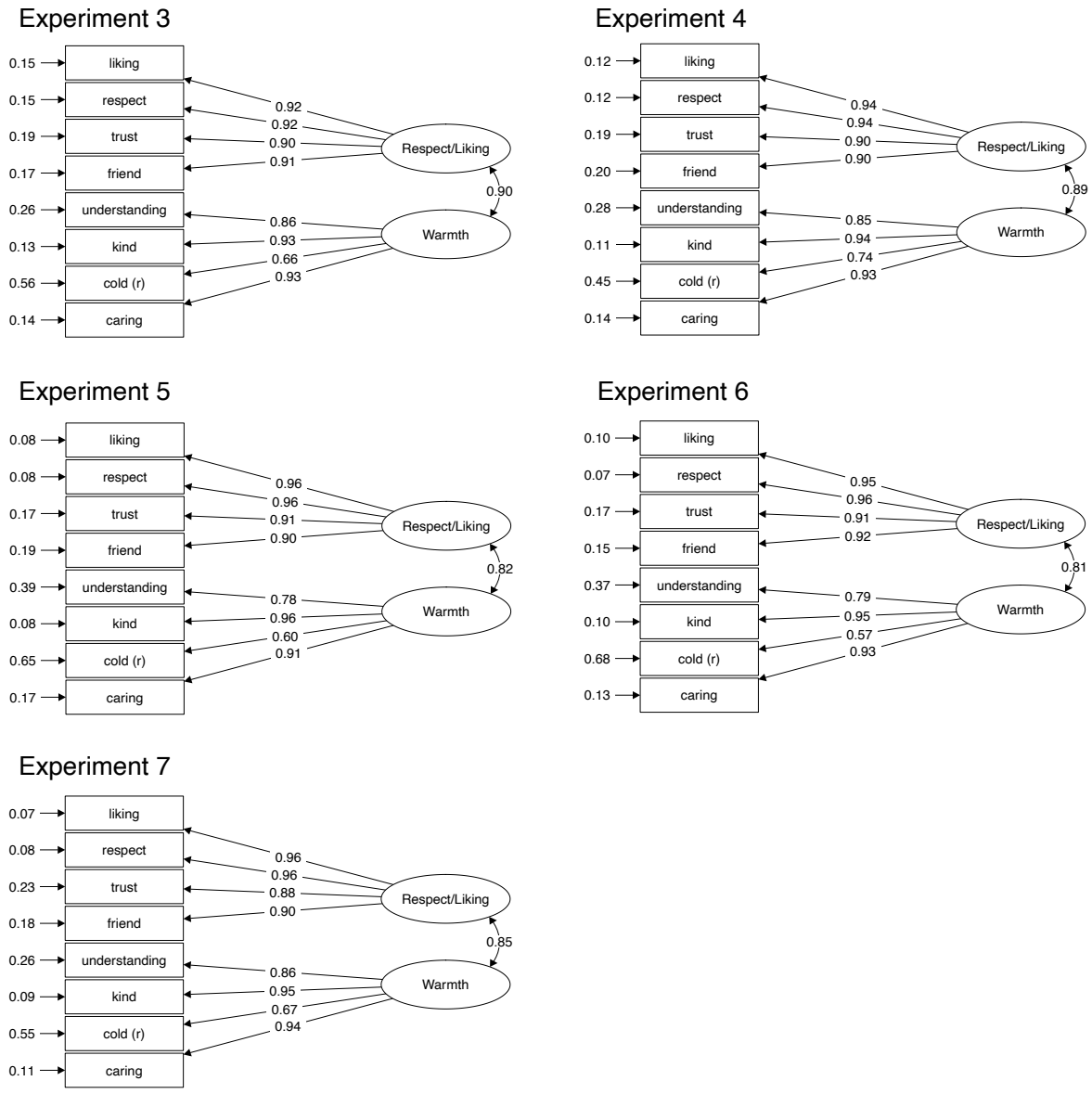

Figure S2. Diagram of the two-factor CFA models in Experiments 3 through 7. We set the variance of each latent variable to 1 in order to identify the scale of the model. All estimates are presented in standardized metric. The item "cold" was reverse-coded.

Full Statistical Models of the Latent Mediation Analyses in Experiments 4 Through 7

We conducted latent moderated mediation analyses in Experiments 4 and 7 and latent simple mediation analyses in Experiments 5 through 7. Diagram of the full statistical model for latent moderated mediation analyses is presented in Figure S3, and diagram of the full statistical model for latent simple mediation analyses is presented in Figure S4.

In Figures S3 and S4, "inferences" are inferences about Beth's attitude toward Ann. For respect/liking, items 1 through 4 indicate how much participants *liked*, *respected*, *trusted*, and *would like to be friends with* Beth; for warmth, items 1 through 4 indicate how *understanding*, *kind*, *cold* (reverse-coded), and *caring* Beth was. The items "like," "positive," and "unfavorable" indicate how much participants agreed that Beth *liked* Ann, felt *positive* toward Ann, and had an *unfavorable* opinion of Ann (reverse-coded). In both models, we allowed the residual covariance between the two positively-worded items of the mediator (agreement with the statements "[Responder] likes [target]" and "[Responder] feels positive toward [target]") to be freely estimated. We did so to reduce model misspecification of the mediator from ignoring wording-related covariance (Marsh, 1996) and better isolate the true mediator variance, which, in turn, should provide greater power and more accurate indirect effect estimates (Gonzalez & MacKinnon, 2020). To retain local independence of the latent mediator, we constrained the factor loadings of those two items to be equal. In the latent moderated mediation model, we additionally allowed the residual covariance between the two product indicators of the Inferences × Target Valence latent variable that involve the two positively-worded items ("Like × Target Valence" and "Positive × Target Valence") to be freely estimated. We constrained the factor loadings of those product indicators to be equal. For visual simplicity, the residual variances of all variables and the covariances of all exogenous variables are omitted from the figures.

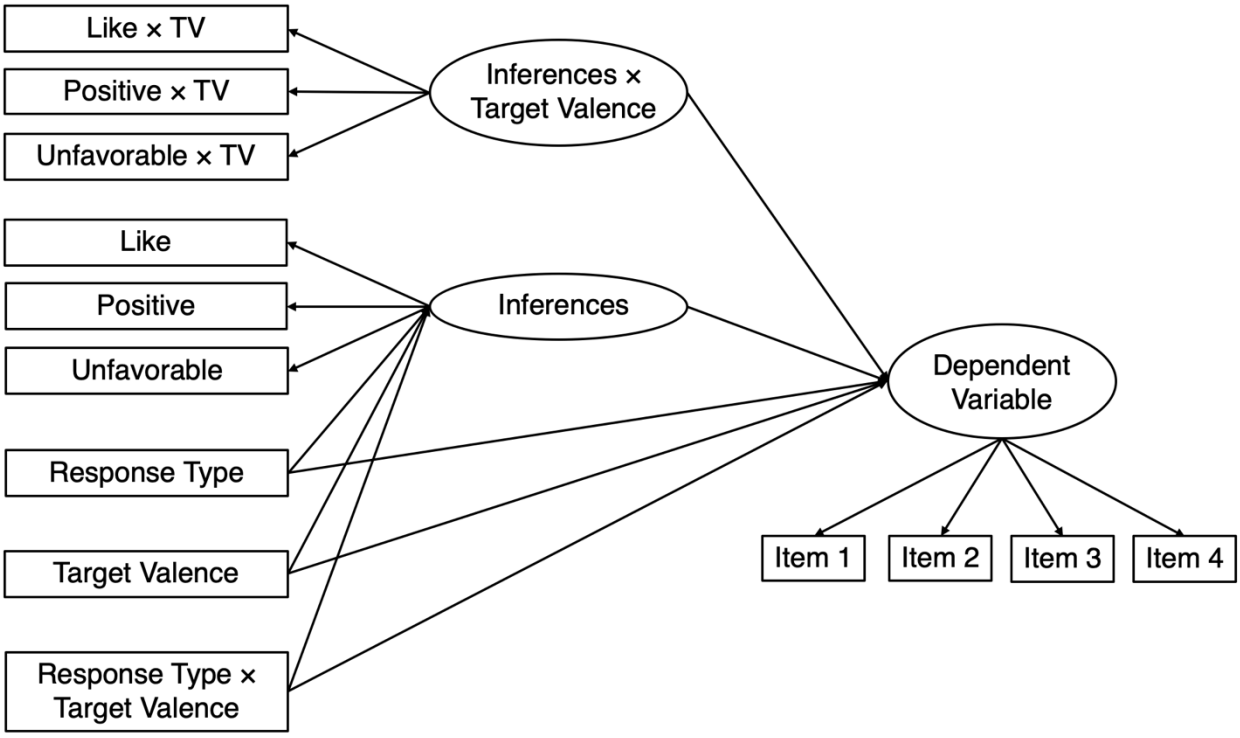

Figure S3. Full statistical model of the latent moderated mediation analyses in Experiments 4 and 7. TV = target valence. The latent interaction term Inferences × Target Valence was measured by the product indicators that were created from the indicators of inferences and the observed target valence variable using the all-pairs approach (Foldnes & Hagvet, 2014; Wall & Amemiya, 2001). Details of the observed predictors (response type, target valence, and Response Type × Target Valence interaction) are reported in the paper.

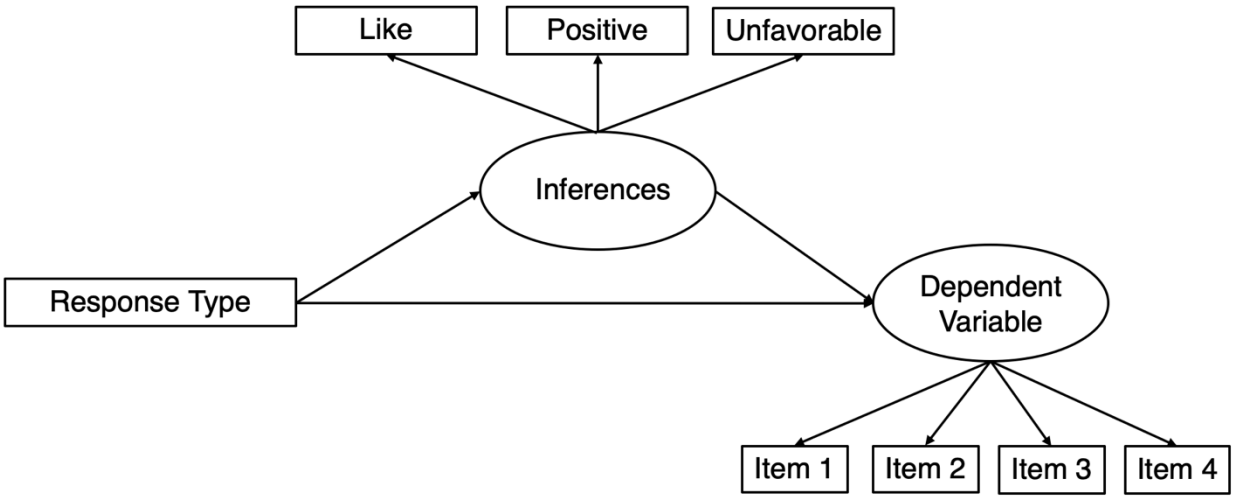

Figure S4. Full statistical model of the latent simple mediation analyses in Experiments 5 through 7. Details of response type are reported in the paper.

Mediational Evidence in Experiments 5 Through 7

Fit of Simple Mediation Models

The fit indices of the latent simple mediation models using our main analytic approach are reported in Table S3.

Table S3

Fit Indices of the Simple Mediation Models in Experiments 5 Through 7 and the Pooled Data.

DV	Dataset	χ^2	CFI	TFI	RMSEA
Respect/liking	Experiment 5	164.62	0.97	0.95	0.13
	Experiment 6	122.60	0.97	0.95	0.12
	Experiment 7	97.29	0.96	0.94	0.14
	Pooled Data	335.56	0.97	0.95	0.13
Warmth	Experiment 5	160.66	0.96	0.93	0.13
	Experiment 6	233.08	0.92	0.87	0.17
	Experiment 7	86.18	0.95	0.93	0.13
	Pooled Data	412.45	0.95	0.92	0.14

Note. Although the fit indices of some models slightly differed from conventional recommendations, inspection of residual matrices suggested that all models fit the data reasonably well and that the fit indices were oversensitive to minor model misspecifications, given the low unique variances of some observed variables (< .10; Browne et al., 2002). In all models, $df = 18$, ps < .001.

Alternative Analytic Approaches to Simple Mediation Models in Experiments 5 Through 7

To assess the robustness of evidence for indirect effects from the simple mediation analyses, we compared the results with those from two alternative analytic approaches. The first analytic approach was almost identical to the main approach but ignored wording differences among the items in the mediator (i.e., the mediator items had freely estimated factor loadings and independent residual variances). This approach reflects our originally intended analytic strategy but resulted in worse model fit across all datasets ($\Delta\chi^2$s = 9.78–42.51). The indirect effect estimates from this first alternative approach had the same level of significance and signs as the estimates from the main analytic approach in Experiments 5 and 7, but not for Experiment 6. The indirect effect from the pooled data was not significant for respect/liking, $ab = 0.07$, $p = .402$,

95% CI [-0.10, 0.24], or warmth, ab = -0.15, p = .090, 95% CI [-0.32, 0.02], with the indirect effect estimate for warmth marginally significant but in the opposite direction as that estimated from the main analytic approach.

The second alternative approach contained only observed (rather than latent) variables and modeled both the mediator and the DVs as composite scores. The indirect effect estimates from this second alternative approach had the same level of significance and signs as the estimates from the main analytic approach for all datasets, including the estimates from the pooled data for both respect/liking, ab = 0.50, 95% CI [0.24, 0.77], and warmth, ab = 0.08, 95% CI [-0.16, 0.32].

Moderated Mediation Models in Experiment 7

We conducted moderated mediation analyses on data from Experiment 7 by entering disclosed experience as a moderator (+1/2 = job stress, -1/2 = cancer stress). The mediation model for respect/liking had acceptable fit, $\chi^2(57) = 772.11$, $p < .001$, CFI = 0.89, TLI = 0.85, RMSEA = 0.16.[7] The Response Type × Disclosed Experience interaction predicted the mediator (a_{mod} = 0.63, p = .007), but the mediator only marginally predicted respect/liking (b = 0.11, p = .064), suggesting no evidence of first-stage moderated mediation ($a_{mod}b$ = 0.07, p = .126). There was no evidence of second-stage moderated mediation either (ab_{mod} = 0.11, p = .526). These results suggest that the effect of response type on inferences about Beth's attitude toward Ann was stronger when Ann disclosed cancer (vs. job) stress (a_{cancer} = -4.10 vs. a_{job} = -3.47), but the associations between inferences about Beth's attitude toward Ann and respect/liking were

[7] We concluded that the latent moderated mediation models in Experiment 7 provided acceptable fit after a holistic assessment of the fit indices as well as the residual matrices and modification indices of these models. Although the fit indices here are less than ideal, we observed low unique variances (< .10) similar to ones observed in Experiment 4 for the majority of the observed variables in these models. These low unique variances suggest that most items were highly reliable (e.g., items on how much participants liked and respected Beth both had standardized factor loadings above .96) and might have led to fit indices that were oversensitive to minor model misfit (Browne et al., 2002). Furthermore, we did not identify any conceptually sensible modification to these models that would non-trivially improve model fit, and none of the key parameter estimates changed in significance when we explored conceptually sensible modifications. Thus, we interpret these models as they were specified in our pre-analysis plan.

comparable across disclosed experience (b_{cancer} = 0.12 vs. b_{job} = 0.09). The overall indirect effect was also comparable across disclosed experience ($a_{cancer}b_{cancer}$ = -0.51 vs. $a_{job}b_{job}$ = -0.32; see Table S4 for all parameter estimates).

Table S4

Parameter Estimates and 95% Confidence Intervals from the Latent Moderated Mediation Models in Experiment 7.

Parameter	Respect/liking	Warmth
a	-3.78 [-4.27, -3.29]	-3.78 [-4.27, -3.29]
a_{job}	-3.47 [-3.98, -2.96]	-3.47 [-3.98, -2.96]
a_{cancer}	-4.10 [-4.67, -3.53]	-4.10 [-4.67, -3.53]
a_{mod}	0.63 [0.18, 1.08]	0.63 [0.18, 1.08]
b	0.11 [-0.01, 0.22]	0.21 [0.09, 0.33]
b_{job}	0.09 [-0.03, 0.22]	0.17 [0.04, 0.30]
b_{cancer}	0.12 [-0.00, 0.25]	0.24 [0.11, 0.37]
b_{mod}	-0.03 [-0.12, 0.06]	-0.07 [-0.17, 0.03]
c	-0.19 [-0.37, 0.00]	-0.94 [-1.14, -0.73]
c'	0.23 [-0.25, 0.70]	-0.16 [-0.65, 0.33]
$a_{job}b_{job}$	-0.32 [-0.76, 0.11]	-0.59 [-1.05, -0.14]
$a_{cancer}b_{cancer}$	-0.51 [-1.02, 0.01]	-0.99 [-1.54, -0.44]

The mediation model for warmth also had acceptable fit, $\chi^2(57) = 806.03$, $p < .001$, CFI = 0.87, TLI = 0.83, RMSEA = 0.17. In addition to the effect of the Response Type × Disclosed Experience interaction on the mediator (a_{mod} = 0.63, p = .006), the mediator predicted warmth (b = 0.21, p = .001), suggesting first-stage moderated mediation ($a_{mod}b$ = 0.13, p = .034). There was no evidence of second-stage moderated mediation (ab_{mod} = 0.26, p = .154). These results suggest that, although the associations between inferences about Beth's attitude toward Ann and warmth were comparable across disclosed experience (b_{cancer} = 0.24 vs. b_{job} = 0.17), the overall indirect effect was stronger when Ann disclosed cancer (vs. job) stress ($a_{cancer}b_{cancer}$ = -0.99 vs. $a_{job}b_{job}$ = -0.59; see Table S4 for all parameter estimates).

Exploratory Analyses on Perceived Similarity Between Responder and Target

In all experiments, we included a single-item exploratory measure on perceived similarity between the responder and the target ("To what extent do you think [responder's name] and [target's name] are similar to each other?" 1 = *not at all*, 7 = *very much*). We conducted a series of analyses to explore the possibility that perceived similarity underlies evaluations of empathizers—namely, that participants in our experiments evaluated empathizers based on how similar they think the empathizer and the target is.

First, we explored whether the Response Type × Target Valence interaction effects that we observed for evaluations of empathizers in Experiments 1 through 4 are present for perceived similarity. Across the experiments, the Response Type × Target Valence ANOVAs showed main effects of response type but no Response Type × Target Valence interaction (except for a small interaction in Experiment 4; see Table S5 for results from each experiment). Given that the Response Type × Target Valence interaction on similarity was largely absent in Experiments 1 through 4, we conclude that similarity is unlikely to have driven evaluations of empathizers in those experiments.

Table S5

ANOVA on perceived similarity in Experiments 1 Through 4.

Experiment	Predictor	df_n	df_d	F	p	η_p^2	$CI_{90\%}$
Experiment 1	Response Type	1	322	66.16	< .001	.17	[.11, .23]
	Target Valence	1	322	12.64	< .001	.04	[.01, .08]
	Interaction	1	322	0.47	.493	< .01	[.00, .02]
Experiment 2	Response Type	1	522	248.34	< .001	.32	[.27, .37]
	Target Valence	1	522	10.88	.001	.02	[.01, .04]
	Interaction	1	522	0.05	.831	< .01	[.00, .00]
Experiment 3	Response Type	1	412	340.78	< .001	.45	[.40, .50]
	Target Valence	1	412	3.28	.071	.01	[.00, .03]
	Interaction	1	412	1.45	.229	< .01	[.00, .02]
Experiment 4	Response Type	2	734	70.69	< .001	.16	[.12, .20]
	Target Valence	1	734	0.03	.853	< .01	[.00, .00]
	Interaction	2	734	5.12	.006	.01	[.00, .03]

Note. df_n = degree of freedom numerator; df_d = degree of freedom denominator; interaction = Response Type × Target Valence interaction.

Next, we explored the indirect effects of similarity. We compared two analytic approaches: With the first approach, we combined similarity with the three items on inferences about the responder's attitude toward the target to form a four-item latent variable (affinity). We then tested for the indirect effects of this latent variable. With the second approach, we tested for the effects of similarity as a single-item mediator in mediation models where all variables were observed. Our mediation analyses focused on Experiments 4 through 7 to compare results from these two approaches with those from our planned analyses (and because the ANOVA results for similarity showed clear divergence from the results for our primary DVs in Experiments 1 through 3).

Estimates of the indirect effects are reported in Figure S5 (Experiment 4) and S6 (Experiments 5 through 7). The first approach yielded results that were highly similar to those from our planned approach (for which the mediator was inferences about the responder's attitude toward the target): Almost all estimates from the first approach had the same signs, levels of significance, and largely overlapping 95% CIs as those from our planned approach. These findings suggested that similarity added little unique variance that accounted for the relations between the predictors and the DVs. Results from the second approach showed a more mixed picture. In almost all models with respect/liking as the DV and some models with warmth as the DV, estimates from the second approach largely matched estimates from the other two approaches in terms of magnitude and signs. In the other models, estimates from the second approach diverged from those from our planned approach, but the direction of divergence differed across models (e.g., the second-stage moderated mediation effect from Model 4 in Experiment 4 was smaller but in the same direction; the mediation effect on warmth in

Experiment 7 was in the opposite direction). Because these differences only appeared in some models and experiments and did not seem consistent, and because using observed versus latent variables tend to yield estimates that are more precise but less accurate (Ledgerwood & Shrout, 2011; Wang & Rhemtulla, in press), we hesitate to draw substantive conclusions from these differences.

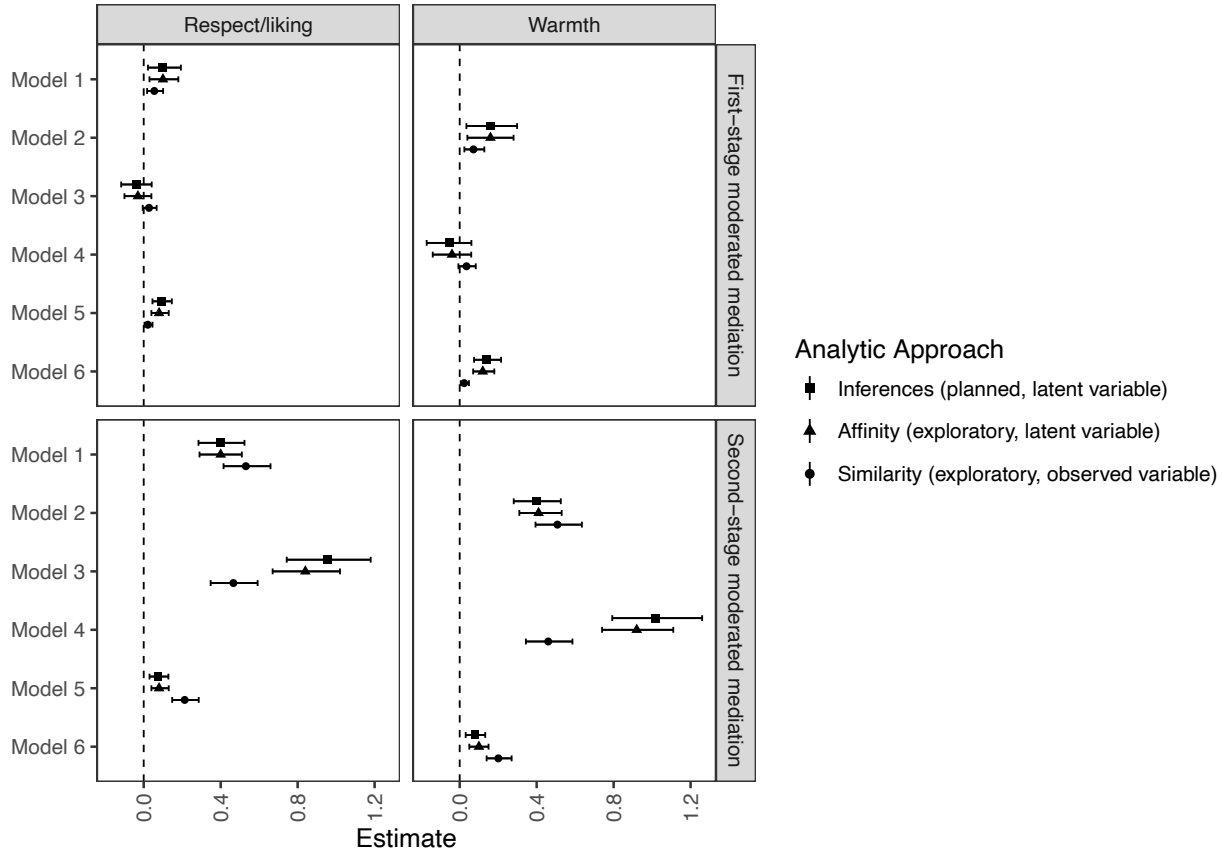

Figure S5. Indirect effect estimates from the moderated mediation analyses in Experiment 4. Respect/liking was the DV in Models 1, 3, and 5; warmth was the DV in Models 2, 4, and 6 (see paper for details on these models). First-stage moderated mediation effect was $a_{mod}b$; second-stage moderated mediation effect was ab_{mod}. Results from the planned approach are also reported in the paper; we reproduce them here for ease of comparison.

Taken together, we speculate that our similarity item functioned like the inferences measure, in that they both seem to capture participants' thoughts about the relationship between the responder and the target (i.e., their affinity). We also note that the meaning of our similarity

measure might have been ambiguous: Participants could have interpreted the item as asking whether the responder and the target had similar backgrounds (e.g., both working for a children's hospital or a White supremacist organization), or more broadly as whether they shared certain values or even demographic characteristics. We believe future research could more fruitfully and rigorously test for similarity as a mechanism by using better measures and manipulating both perceived similarity and inference affordance.

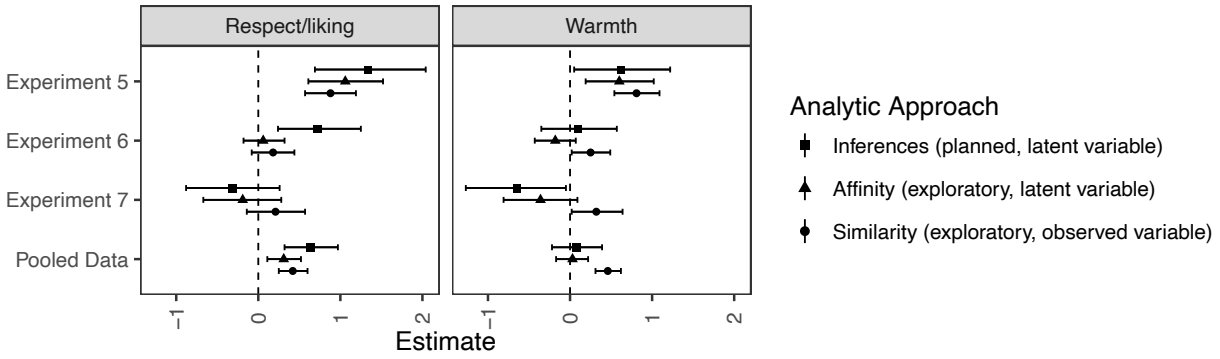

Figure S6. Indirect effect (*ab*) estimates from simple mediation analyses in Experiments 5 through 7 and the pooled data (see paper for details on these datasets). Results from the planned approach are also reported in the paper; we reproduce them here for ease of comparison.

References

Browne, M. W., MacCallum, R. C., Kim, C. T., Andersen, B. L., & Glaser, R. (2002). When fit indices and residuals are incompatible. *Psychological Methods*, *7*, 403-421.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175-191.

Foldnes, N., & Hagtvet, K. A. (2014). The choice of product indicators in latent variable interaction models: Post hoc analyses. *Psychological Methods*, *19*, 444-457.

Giner-Sorolla, R. (2018). Powering your interaction. Retrieved from https://approachingblog.wordpress.com/2018/01/24/powering-your-interaction-2/

Gonzalez, O., & MacKinnon, D. P. (2020). The measurement of the mediator and its influence on statistical mediation conclusions. *Psychological Methods*. Advanced online publication.

Ledgerwood, A. (2019). New developments in research methods. In E. J. Finkel & R. F. Baumeister (Eds.), *Advanced social psychology* (pp. 39-61). Oxford University Press.

Ledgerwood, A., & Shrout, P. E. (2011). The trade-off between accuracy and precision in latent variable models of mediation processes. *Journal of Personality and Social Psychology, 101*, 1174-1188.

Marsh, H. W. (1996). Positive and negative global self-esteem: A substantively meaningful distinction or artifactors? *Journal of Personality and Social Psychology*, *70*, 810–819.

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, *48*, 1-36.

Schoemann, A. M., Boulton, A. J., & Short, S. D. (2017). Determining power and sample size for simple and complex mediation models. *Social Psychological and Personality Science*, *8*, 379-386.

Wall, M. M., & Amemiya, Y. (2001). Generalized appended product indicator procedure for nonlinear structural equation analysis. *Journal of Educational and Behavioral Statistics*, *26*, 1-30.

Wang, Y. A., & Rhemtulla, M. (in press). Power analysis for parameter estimation in structural equation modeling: A discussion and tutorial. *Advances in Methods and Practices in Psychological Science*.