Supplementary Materials

This document provides Supplementary Materials for the manuscript "The Empathetic Refutational Interview to tackle vaccine misconceptions: four randomised experiments". Pre-registrations, experimental materials, data, and the code used to derive the analyses reported herein are shared on the Open Science Framework (OSF): https://osf.io/knq9g

Wording of all questions used in experiments

Note that all of these items are also provided on the OSF at the link above, including replication-ready formats that can be implemented directly in the survey software used. Instructions for questions are provided here, followed by the question items and their response scales. Words in italics were not shown to participants.

Vaccine acceptance

5C scale (Betsch et al., 2018)

These questions are about vaccination. Please indicate how much you agree with the statements below.

Confidence subscale

- I am completely confident that vaccines are safe.
- Vaccinations are effective
- Regarding vaccines, I am confident that public authorities decide in the best interest of the community.

Complacency subscale

- Vaccination is unnecessary because vaccine-preventable diseases are not common anymore. (R)
- My immune system is so strong, it also protects me against diseases. (R)
- Vaccine-preventable diseases are not so severe that I should get vaccinated. (R)

Collective responsibility subscale

- When everyone is vaccinated, I don't have to get vaccinated, too. (R)
- I get vaccinated because I can also protect people with a weaker immune system.
- Vaccination is a collective action to prevent the spread of diseases.

Response options: Likert scale, 1: strongly disagree – 7: strongly agree (R) denotes reversed-coded item such that higher scores reflect higher vaccine acceptance.

COVID-19 vaccine willingness question

If a COVID-19 booster vaccine is recommended yearly, how willing are you to get it?

Response items: Likert scale from not at all willing – very willing. This was a 5-point scale in Experiment 3 and a 7-point scale in Experiment 4.

Argument ratings

You will now see some reasons why people may not wish to be vaccinated.

Here, arguments are shown to participants; 10 in Experiment 1, 6 in Experiments 3 and 4 due to time constraints; full list of arguments is available on the OSF.

1. Please rate your level of support for the position expressed in the statement.

Response options: Likert scale, 1: strongly against, 2: moderately against, 3: slightly against, 4: neutral, 5: slightly in favour, 6: moderately in favour, 7: strongly in favour

For top selected arguments, each was then shown with the following questions:

2. How well would you say you understand the position expressed in this statement?

Response options: Likert scale, 1: not at all– 7: completely

3. How well do you think you are able to justify the position expressed in this statement?

Response options: Likert scale, 1: very poorly–7: very well

Instructions for soliciting explanations vs. control (Experiment 1)

This procedure was taken exactly from Fernbach et al. (2013)

Control condition:

> Please write down all the reasons you have for your position on this statement, going from the most important to the least. That is, you should state precisely why you hold this position. Try to tell as complete a story as you can about the reasons for your position.
>
> Please take your time, as we expect you to carefully state your reasons.

Experimental (soliciting explanations) condition:

> Please describe all the details you know about why you have this position for the statement, going from the first step to the last, and providing the causal connection between these details. That is, your explanation should state precisely how each detail links to the others, in one continuous chain from start to finish. In other words, try to tell as complete a story as you can, with no gaps.
>
> Please take your time, as we expect your best explanation.

Ratings of HCP-patient scenario (Experiments 2-4)

Participants read interactions between Dr Jones and Tom, as described in the main manuscript. A full table of all scenarios is provided on the OSF. For each interaction, participants were asked the following questions.

Experiment 2 only

These questions are about **Tom's** post.

1. How much do you support **Tom's** post?

Response options: Likert scale, 1: strongly against, 7: strongly in favour

2. Irrespective of whether you support it, how compelling do you think **Tom's** post is?

Response options: Likert scale, 1: not at all, 7: completely

3. Irrespective of whether you support it, suppose someone asks you to justify **Tom's** claim. How compelling do you think your justification would be?

Response options: Likert scale, 1: not at all, 7: completely

These questions are about **Dr Jones'** post.

1. How much do you support **Dr Jones'** post?

Response options: Likert scale, 1: strongly against, 7: strongly in favour

2. Irrespective of whether you support it, how compelling do you think **Dr Jones'** post is?

Response options: Likert scale, 1: not at all, 7: completely

3. Irrespective of whether you support it, suppose someone asks you to justify **Dr Jones'** claim. How compelling do you think your justification would be?

Response options: Likert scale, 1: not at all, 7: completely

Experiment 3 and 4

1. How much do you support Dr Jones' response to Tom?

Response options: Likert scale, 1: strongly against –7: strongly in favour

2. Irrespective of whether you support it, how compelling do you think Dr Jones' response to Tom is?

Response options: Likert scale, 1: not at all – 7: completely

(Scenario continues before next two questions.

3. Regardless of whether you supported Tom's statements, how open are you to hearing more from Dr Jones?

Response options: Likert scale, 1: not at all – 7: completely

4. Adapted Trust in a Medical Doctor scale (Dugan et al., 2005): Thinking about the overall interaction between Dr Jones and Tom, please rate your agreement with each of the following statements.

- Dr Jones cares more about what is convenient than about Tom's medical needs. (R)
- Dr Jones is extremely thorough and careful.
- Tom can completely trust Dr Jones' decisions about whether vaccination is best for him.
- Dr Jones is totally honest in telling Tom about the vaccination.
- All in all, Tom can have complete trust in Dr Jones.

Response options: Likert scale, 1: Strongly disagree – 5 Strongly agree

Covariates

Trust in the Medical Profession scale (Dugan et al., 2005)

Please rate your agreement with each of the following statements.

- Sometimes doctors care more about what is convenient for them than about their patients' medical needs. [reverse-coded variable]

- Doctors are extremely thorough and careful.
- I completely trust doctors' decisions about which medical treatments are best.
- A doctor would never mislead me about anything.
- All in all, I trust doctors completely.

Response options: Likert scale, 1: Strongly disagree – 5 Strongly agree

6-item Cognitive Reflection Test (Experiment 2 only Primi et al., 2016)

Please write your answer to the following questions in the text box provided.

1. A bat and a ball cost £1.10 in total. The bat costs £1.00 more than the ball. How much does the ball cost? (Please give the amount in £)

2. If it takes 5 minutes for five machines to make five widgets, how long would it take for 100 machines to make 100 widgets? (Please give the number of minutes)

3. In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake? (Please give the number of days)

4. If three elves can wrap three toys in an hour, how many elves are needed to wrap six toys in 2 hours? (Please give the number of elves)

5. Jerry received both the 15th highest and the 15th lowest mark in the class. How many students are there in the class? (Please give the number of students)

6. In an athletics team, tall members are three times more likely to win a medal than short members. This year the team has won 60 medals so far. How many of these have been won by short athletes? (Please give the number of medals)

The mean score on the CRT in Experiment 2 was 3.12, SD = 1.93.

Health Literacy Scale (Experiment 2 only Haun et al., 2009)

Please answer the following questions:

- How often do you have someone help you read hospital materials? (R)
- How often do you have problems learning about your medical condition because of difficulty understanding written information? (R)
- How often do you have a problem understanding what is told to you about a medical condition? (R)
- How confident are you filling out medical forms by yourself?

The mean score for the Health Literacy Scale in Experiment 2 was 4.40, SD = 0.62, Cronbach's $\alpha = .76$.

Response options: Likert scale, 1: never, 2: occasionally, 3: sometimes, 4: often, 5: always

Detailed results of statistical tests, data visualisation, and exploratory analyses

Table S1

Results of between-within ANOVAs on argument ratings in Experiment 1

Effect	$F(1, 224)$	p-value	η_P^2
Support for argument			
Experimental condition	0.51	.476	<0.01
Timing (pre/post test)	34.62	< .001	0.13
Interaction	0.82	.367	<0.01
Understanding of argument			
Experimental condition	0.17	.683	<0.01
Timing (pre/post test)	16.97	< .001	0.07
Interaction	1.75	.187	0.01
Ability to justify position			
Experimental condition	2.87	.092	0.01
Timing (pre/post test)	3.60	.059	0.02
Interaction	5.42	.021	0.02

Table S2

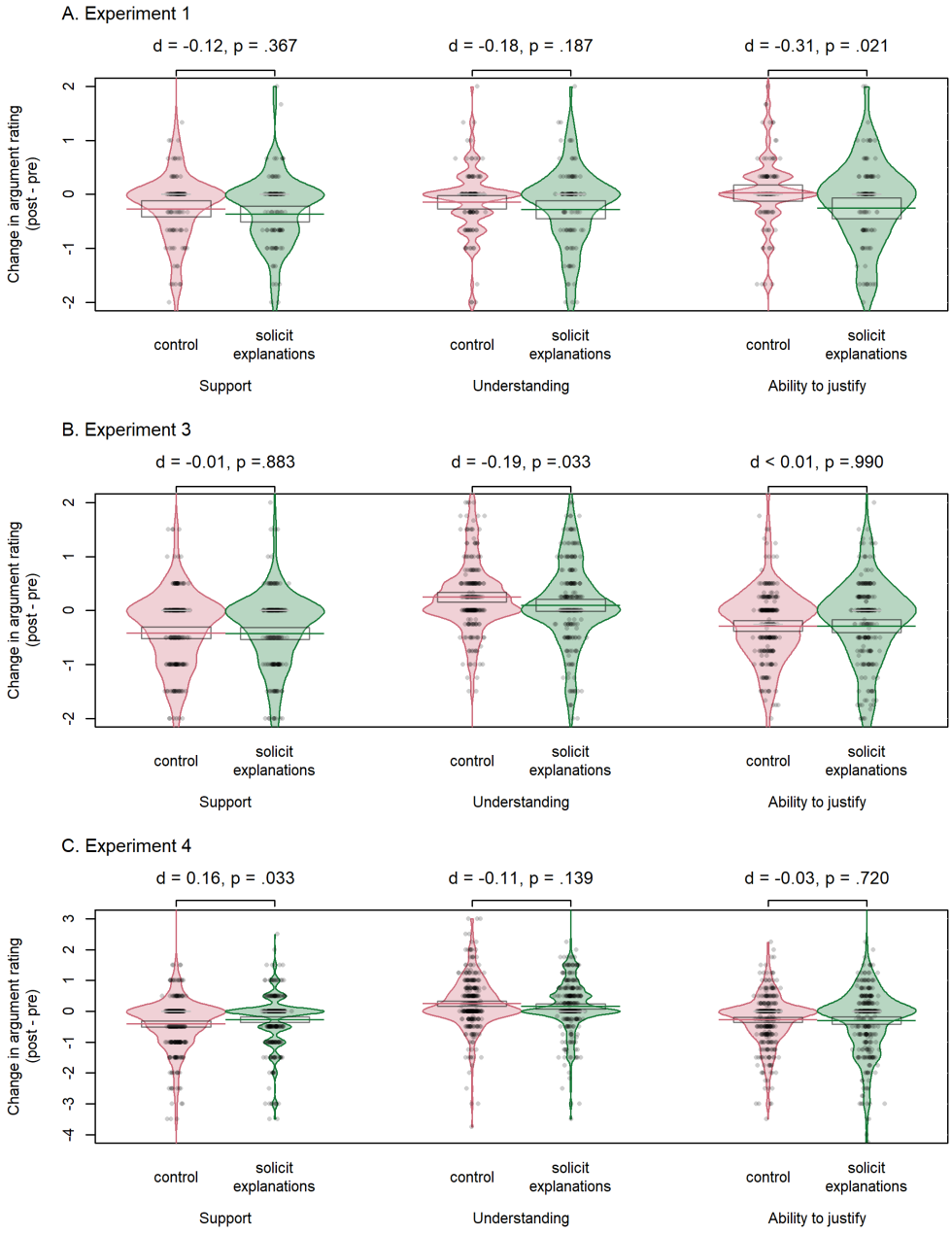
Pre- and post-test vaccine acceptance scores in Experiment 1

	Cronbach's α		Control (n = 115) Mean (SD)		Solicit Explanations (n = 111) Mean (SD)	
	Pre-test	Post-test	Pre-test	Post-test	Pre-test	Post-test
Confidence	.87	.87	3.46 (1.67)	3.37 (1.57)	3.57 (1.65)	3.48 (1.62)
Complacency	.72	.85	4.07 (1.28)	4.08 (1.45)	3.79 (1.35)	3.93 (1.44)
Collective responsibility	.63	.66	4.09 (1.39)	4.07 (1.31)	4.06 (1.43)	4.08 (1.47)
Between-within ANOVA results						
Effect	$F(1, 224)$		p-value		η_P^2	
Confidence						
Experimental condition	0.28		.598		<0.01	
Timing (pre/post test)	2.83		.094		0.01	
Interaction	0.01		.932		<0.01	
Complacency						
Experimental condition	1.46		.228		<0.01	
Timing (pre/post test)	2.15		.144		0.01	
Interaction	1.53		.217		0.01	
Collective responsibility						
Experimental condition	<0.01		.953		<0.01	
Timing (pre/post test)	<0.01		.979		<0.01	
Interaction	0.12		.725		<0.01	

Note. Because the collective responsibility subscale had less-than-satisfactory reliability (< .70), we checked whether the analyses would change using a two-item composite rather than a three-item (dropping the item that resulted in reliability of α = .75 on both subscales). Using either composite measure did not change the outcome of any of the analyses.

Figure S1

Changes in anti-vaccination argument ratings in the control and soliciting-explanations conditions across Experiments 1, 3, and 4



Note. Violins and black dots show the distribution of responses, with the means given by coloured lines (grey boxes show 95% confidence intervals). Cohen's *d* effect sizes and *p*-values were calculated as the difference in difference between conditions.

Table S3

Results of ANCOVAs on ratings of Dr Jones in Experiment 2

Effect	$F(1, 1094)$	p-value	η_P^2	$M_{control}$ (*SD*)	$M_{affirming}$ (*SD*)
Support for Dr Jones' refutation					
Experimental condition	65.92	$< .001$	0.06	4.23 (1.80)	4.93 (1.59)
Baseline argument rating	212.29	$< .001$	0.16	4.08 (1.87)	4.12 (1.80)
Trust in medical profession	183.47	$< .001$	0.14		
Health literacy	1.87	.172	< 0.01		
CRT score	1.12	.290	< 0.01		
How compelling was Dr Jones' refutation					
Experimental condition	261.35	$< .001$	0.19	3.72 (1.70)	5.18 (1.46)
Baseline argument rating	25.74	$< .001$	0.02	4.21 (1.77)	4.32 (1.73)
Trust in medical profession	110.73	$< .001$	0.09		
Health literacy	2.63	.105	< 0.01		
CRT score	2.14	.144	< 0.01		
Ability to justify Dr Jones' refutation					
Experimental condition	43.41	$< .001$	0.04	4.26 (1.63)	4.87 (1.47)
Baseline argument rating	11.67	$< .001$	0.01	4.21 (1.69)	4.13 (1.63)
Trust in medical profession	111.18	$< .001$	0.09		
Health literacy	0.27	.606	< 0.01		
CRT score	4.82	.028	< 0.01		
Support for Tom's (anti-vax) position					
Experimental condition	0.06	.809	< 0.01	3.42 (1.88)	3.45 (1.85)
Baseline argument rating	382.11	$< .001$	0.25	4.08 (1.87)	4.12 (1.80)
Trust in medical profession	75.70	$< .001$	0.07		
Health literacy	0.23	.632	< 0.01		
CRT score	9.87	.002	0.01		
How compelling was Tom's (anti-vax) position					
Experimental condition	1.06	.303	< 0.01	3.39 (1.81)	3.33 (1.80)
Baseline argument rating	170.14	$< .001$	0.14	4.21 (1.77)	4.32 (1.73)
Trust in medical profession	40.53	$< .001$	0.04		
Health literacy	0.25	.619	< 0.01		
CRT score	15.60	$< .001$	0.01		
Ability to justify Tom's (anti-vax) position					
Experimental condition	0.01	.933	< 0.01	3.58 (1.77)	3.51 (1.73)
Baseline argument rating	330.10	$< .001$	0.23	4.21 (1.69)	4.13 (1.63)
Trust in medical profession	39.97	$< .001$	0.04		
Health literacy	0.14	.708	< 0.01		
CRT score	3.86	.050	< 0.01		

Note. Mean (SD) trust in medical profession score was 2.92 (0.84) in the control and 2.98 (0.82) in the affirming condition. Mean (SD) health literacy was 4.40 (0.64) in the control and 4.40 (0.61) in the affirming condition. Mean (SD) CRT score was 3.25 (1.91) in the control and 3.00 (1.95) in the affirming condition.

Exploratory subgroup analyses in Experiment 2

In Experiment 2, to further assess the interplay between existing attitudes and support for the HCP's refutation, we conducted an exploratory subgroup analysis based on participants' opinions of the COVID-19 vaccine, as we had in our sample three different opinion groups: positive, negative, or neutral/not stated. Participants with positive opinions showed the highest support for the refutation, followed by those who were neutral, and finally those with negative opinions (see Figure S2). A between-subjects ANCOVA including this factor and its interaction with experimental condition (while controlling for 22 baseline argument support and trust as covariates) found a significant interaction, $F(2, 1092) = 3.76$, $p = .024$, $\eta_P^2 = 0.01$. Follow-up tests showed that the intervention produced larger effects for those with less positive opinions of the COVID-19 vaccine (negative, $t(399) = 4.87$, $p < .001$, $d = 0.49$; neutral/not stated, $t(454) = 4.84$, $p < .001$, $d = 0.45$; positive, $t(173) = 1.00$, $p = .318$, $d = 0.14$).

Figure S2

Ratings of the HCP's response in the control and experimental conditions in Experiment 2, by participants' attitude towards the COVID-19 vaccines

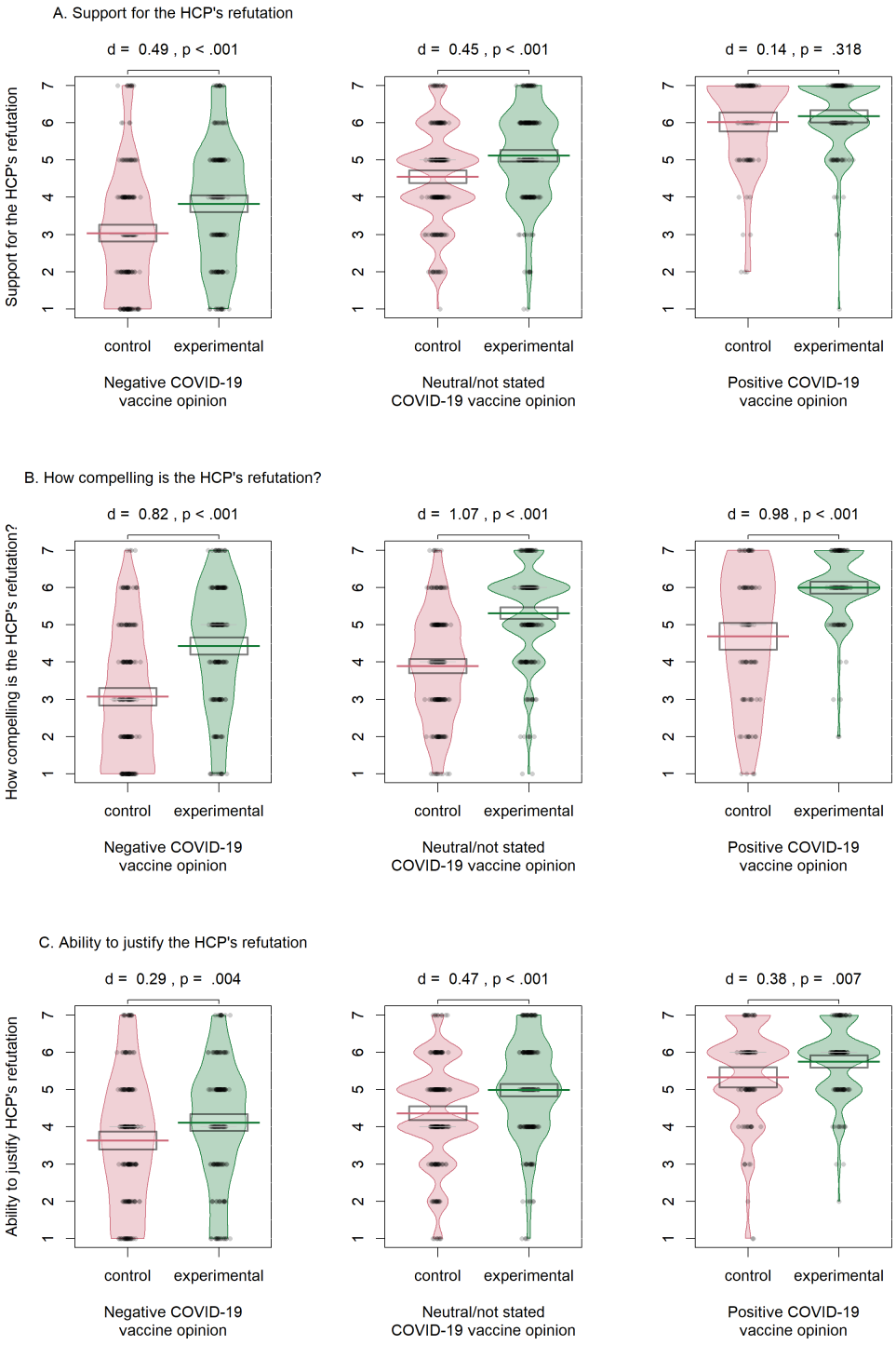


Table S4

Pre- and post-test vaccine acceptance scores in Experiment 2

	Cronbach's α		Control (n = 521) Mean (SD)		Affirming refutation (n = 579) Mean (SD)	
	Pre-test	Post-test	Pre-test	Post-test	Pre-test	Post-test
Confidence	.90	.91	4.34 (1.72)	4.27 (1.72)	4.51 (1.67)	4.47 (1.64)
Complacency	.78	.81	3.20 (1.39)	3.17 (1.42)	3.24 (1.41)	3.17 (1.45)
Between-within ANOVA results						
Effect	F(1, 1098)		p-value		η_P^2	
Confidence						
Experimental condition	3.53		.061		<0.01	
Timing (pre/post test)	9.59		.002		0.01	
Interaction	1.23		.268		<0.01	
Complacency						
Experimental condition	0.07		.796		<0.01	
Timing (pre/post test)	7.25		.007		0.01	
Interaction	1.26		.263		<0.01	

Table S5

Results of between-within ANOVAs on argument ratings in Experiment 3

Effect	F(1, 517)	p-value	η_P^2
Support for argument			
Experimental condition	0.01	.906	<0.01
Timing (pre/post test)	118.55	< .001	0.19
Interaction	0.02	.883	<0.01
Understanding of argument			
Experimental condition	< 0.01	.964	<0.01
Timing (pre/post test)	23.49	< .001	0.04
Interaction	4.59	.033	0.01
Ability to justify position			
Experimental condition	0.557	.457	0.01
Timing (pre/post test)	54.33	< .001	0.02
Interaction	< 0.01	.990	< 0.01
Further changes in argument support after step 1			
Experimental condition	< 0.01	.984	< 0.01
Timing (post-step 1/end of experiment)	22.95	< .001	0.04
Interaction	0.02	.902	< 0.01

Table S6

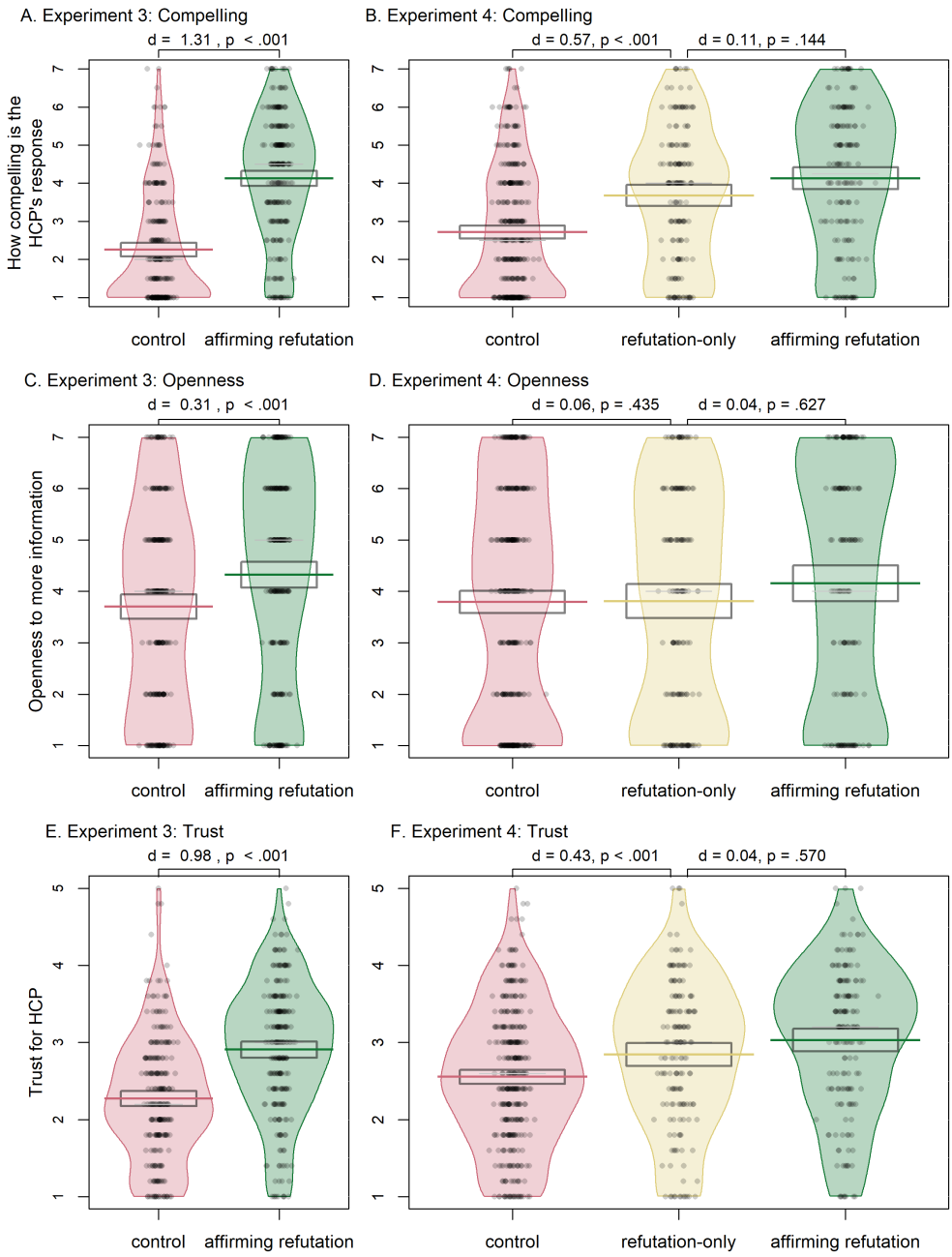
Results of ANCOVAs on ratings of Dr Jones in Experiment 3

Effect	$F(1, 515)$	p-value	η_P^2	Control M (SD)	Affirming M (SD)
Support for Dr Jones' refutation					
Experimental condition	69.27	$< .001$	0.12	2.54 (1.51)	3.45 (1.69)
Baseline argument rating	157.22	$< .001$	0.23		
Trust in medical profession	84.45	$< .001$	0.14		
How compelling was Dr Jones' refutation					
Experimental condition	221.96	$< .001$	0.30	2.26 (1.44)	4.13 (1.63)
Baseline argument rating	26.72	$< .001$	0.05		
Trust in medical profession	33.62	$< .001$	0.06		
Openness to Dr Jones' information					
Experimental condition	12.61	$< .001$	0.02	3.70 (1.94)	4.32 (2.05)
Baseline argument rating	10.71	.001	0.02		
Trust in medical profession	27.22	$< .001$	0.05		
Trust in Dr Jones					
Experimental condition	122.74	$< .001$	0.19	2.27 (0.79)	2.91 (0.85)
Baseline argument rating	58.59	$< .001$	0.10		
Trust in medical profession	168.02	$< .001$	0.25		

Note. Mean (SD) baseline argument rating was 6.15 (1.13) in the control and 6.17 (1.15) in the affirming condition. Mean (SD) trust in medical profession score was 2.56 (0.81) in the control and 2.62 (0.77) in the affirming condition.

Figure S3

Ratings of compellingness, trust, and openness for the HCP's response in the control and experimental conditions in Experiments 3 (left panels) and 4 (right panels)



Note. Violins and black dots show the distribution of responses, coloured lines show the mean and grey boxes show 95% confidence intervals. Cohen's *d* effect sizes and *p*-values were calculated using models controlling for participants' anti-vaccination argument support and trust in the medical profession.

Table S7

Pre- and post-test vaccine acceptance scores in Experiment 3

	Cronbach's α		Control (n = 263) Mean (SD)		ERI (n = 256) Mean (SD)	
	Pre-test	Post-test	Pre-test	Post-test	Pre-test	Post-test
Composite	.88	.88	2.87 (1.23)	3.03 (1.26)	2.77 (1.27)	3.01 (1.34)
Confidence	.88	.89	2.54 (1.49)	2.69 (1.49)	2.55 (1.55)	2.78 (1.60)
Complacency	.71	.77	4.27 (1.31)	4.10 (1.41)	4.44 (1.40)	4.22 (1.49)
Collective	.66	.70	3.42 (1.41)	3.78 (1.38)	3.38 (1.50)	3.59 (1.47)
Booster willingness	-	-	1.47 (1.01)	1.51 (1.06)	1.41 (1.00)	1.52 (1.11)
Between-within ANOVA results						
Effect	$F(1, 517)$		p-value		η_P^2	
Composite measure						
Experimental condition	0.32		.570		<0.01	
Timing (pre/post test)	104.69		< .001		0.17	
Interaction	3.98		.046		0.01	
Confidence						
Experimental condition	0.13		.715		<0.01	
Timing (pre/post test)	44.02		< .001		0.08	
Interaction	2.48		.116		0.01	
Complacency						
Experimental condition	1.59		.208		<0.01	
Timing (pre/post test)	35.56		< .001		0.06	
Interaction	0.60		.439		<0.01	
Collective						
Experimental condition	0.83		.363		<0.01	
Timing (pre/post test)	70.84		< .001		0.12	
Interaction	0.71		.400		<0.01	
Booster willingness						
Experimental condition	0.08		.774		<0.01	
Timing (pre/post test)	11.16		< .001		0.02	
Interaction	2.62		.106		0.01	

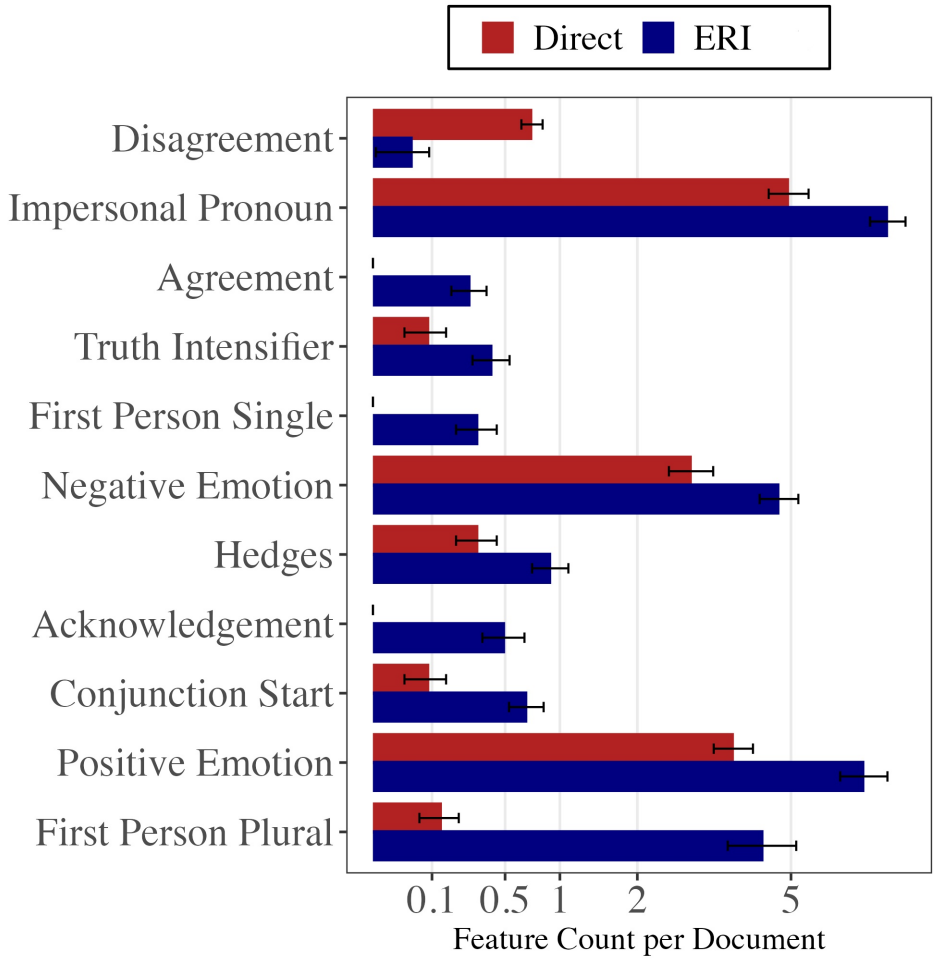
Note. Because the collective responsibility subscale had less-than-satisfactory reliability ($< .70$), we checked whether the analyses would change using a two-item composite rather than a three-item (dropping the item that would result in a reliability of $\alpha > .70$ on both subscales). Using either composite measure did not change the outcome of any of the analyses.

Design and test of HCP's responses for ERI and refutation-only conditions

We crafted two versions of the HCP-patient scenario for the refutation-only and ERI conditions (Table S8) for Experiment 4. To test that these conditions differed in their affirmation and tone, we conducted an analysis using the politeness package in R (Yeomans et al., 2018) to identify differences in the linguistic markers that indicate a speaker's openness towards and willingness to engage with their interlocutor even in situations where they may disagree (Yeomans et al., 2020). Such linguistic markers have also been found to be perceived as more empathetic (Minson et al., 2023). Figure S4 shows the significant differences in these markers between the two conditions. In this case, we are concerned most with comparisons in markers of disagreement (more observed in the direct refutations written without affirmation or empathetic language), agreement (more observed in the ERI texts), hedges (typically higher for more receptive texts, as is the case with the ERI texts here), acknowledgement (more observed in the ERI texts), and emotions. Note that for emotions, both negative and positive emotions are greater in the ERI texts, but this is actually consistent with features in known "warm" and "tough" negotiation texts provided as comparisons by Yeomans et al. (2018), as the warm (receptive/polite) texts also display more of both emotions.

Figure S4

Comparison of linguistic markers using the politeness package in R



Note. Plot shows the average feature count per document in square root scale, including only features that vary meaningfully across the conditions.

Table S8

Example of an anti-vaccination argument and HCP's response in the control, affirming refutation, and refutation-only conditions used in Experiments 2, 3, & 4.

Tom: I don't think people should be getting vaccinated. Medical authorities are overreacting, with vaccines being recommended for every minor illness now.		
Control	Affirming refutation	Refutation-only
Dr Jones: I know that the COVID-19 vaccines are safe and effective.	**Dr Jones:** I can see why you wouldn't automatically trust that vaccines are necessary. We definitely don't want to overuse any type of medicine. It's true that sometimes diseases can seem mild. But it's impossible to know ahead of time how you'll be affected by a disease. We select recommended vaccines because they give good protection against diseases that can have severe consequences. For example, vaccines have successfully reduced the severity of COVID-19 to the extent that a vaccinated person can now experience a mild infection, when before it may have been life-threatening for them.	**Dr Jones:** That's not true. It's impossible to know ahead of time how one will be affected by a disease. The recommended vaccines are selected because they give good protection against diseases that can have severe consequences. For example, vaccines have successfully reduced the severity of COVID-19 to the extent that a vaccinated person can now experience a mild infection, when before it may have been life-threatening for them.

Note. This is one argument drawn from a larger pool, as described in the text. The full set of arguments and refutations are available on the OSF. In Experiment 2, the type of vaccine was not specified, but "COVID-19 vaccine" as shown in this example was used in Experiments 3 and 4. The refutation-only condition was used only in Experiment 4. This condition used the same rebuttal of the argument as the affirming refutation, but made two critical adjustments: (1) It omitted the affirmation, replacing it with the statement "That's not true"; (2) It used colder and more passive language so it would come across as less empathetic than the affirming refutation. These adjustments allowed us to test the incremental effect of being affirming and empathetic in a refutation.

Table S9

Results of mixed ANOVAs on argument ratings in Experiment 4

Effect	$F(1, 696)$	p-value	η_P^2
Support for argument			
Experimental condition	1.42	.234	<0.01
Timing (pre/post test)	102.55	< .001	0.13
Interaction	4.57	.033	0.01
Understanding of argument			
Experimental condition	0.12	.726	<0.01
Timing (pre/post test)	44.06	< .001	0.06
Interaction	2.19	.139	<0.01
Ability to justify position			
Experimental condition	0.35	.556	<0.01
Timing (pre/post test)	65.66	< .001	0.09
Interaction	0.13	.720	<0.01
Further change in argument support after step 1			
Experimental condition	1.06	.367	0.01
Timing (post-step 1/end of experiment)	26.34	< .001	0.04
Interaction	8.21	< .001	0.03

Table S10

Results of ANCOVAs on ratings of Dr Jones in Experiment 4

Effect	F	p-value	η_P^2
Support for Dr Jones' refutation			
Experimental condition	13.02	< .001	0.04
Baseline argument rating	166.86	< .001	0.19
Trust in medical profession	101.50	< .001	0.13
How compelling was Dr Jones' refutation			
Experimental condition	55.41	< .001	0.14
Baseline argument rating	55.84	< .001	0.07
Trust in medical profession	96.27	< .001	0.12
Experimental condition	1.00	.370	< 0.01
Baseline argument rating	60.46	< .001	0.08
Trust in medical profession	60.94	< .001	0.08
Trust in Dr Jones			
Experimental condition	27.63	< .001	0.07
Baseline argument rating	104.73	< .001	0.13
Trust in medical profession	291.70	< .001	0.30

Note. Degrees of freedom was 2, 695 for experimental effects and 1, 695 for covariates.

Table S11

Results of key contrasts for ratings of Dr Jones in Experiment 4

Rating	*M* (*SD*)			Contrast	
	Control	Refutation-only	Affirming	Control vs. both refutations	Refutation vs. affirming
Support	3.05 (1.72)	3.19 (1.67)	3.83 (1.88)	$t = 4.44$, $p < .001$	$t = 2.48$, $p = .013$
Compellingness	2.72 (1.65)	3.68 (1.78)	4.13 (1.87)	$t = 10.40$, $p < .001$	$t = 1.46$, $p = .144$
Openness	3.79 (2.14)	3.81 (2.13)	4.16 (2.29)	$t = 1.33$, $p = .185$	$t = 0.49$, $p = .627$
Trust	2.56 (0.89)	2.84 (0.96)	3.03 (0.96)	$t = 7.42$, $p < .001$	$t = 0.57$, $p = .570$

Note. Degrees of freedom for *t*-values of the contrasts were 696 for control vs. refutations and 695 for refutation vs. affirming conditions.

Table S12

Pre- and post-test vaccine acceptance scores in Experiment 4

Measure	Cronbach's α		Mean pre/post-test change (95% CI)			
	Pre-test	Post-test	Control ($n = 185$)	Facts-only $n = 185$)	Refutation-only ($n = 164$)	ERI ($n = 166$)
Composite	.89	.89	0.02 (-0.04, 0.08)	0.15 (0.08, 0.22)	0.09 (0.03, 0.16)	0.13 (0.07, 0.19)
Confidence	.89	.89	-0.06 (-0.15, -0.04)	0.05 (-0.07, 0.16)	0.10 (-0.01, 0.21)	0.12 (0.02, 0.23)
Complacency	.71	.78	-0.17 (-0.29, -0.05)	-0.19 (-0.30, -0.09)	-0.09 (-0.20, 0.03)	-0.13 (-0.24, -0.03)
Collective	.75	.76	0.04 (-0.07, 0.15)	0.18 (0.08, 0.29)	0.12 (0.01, 0.22)	0.16 (0.05, 0.28)
Booster willingness	-	-	-0.07 (-0.17, 0.03)	0.18 (0.07, 0.30)	0.07 (-0.05, 0.19)	0.11 (-0.01, 0.22)
ANOVA results						
Effect	df	F	*p*-value	η_P^2		
Composite						
Experimental condition	3, 696	0.76	.517	< 0.01		
Timing (pre/post test)	1, 696	36.20	$< .001$	0.05		
Interaction	3, 969	3.16	.024	0.01		
Confidence						
Experimental condition	3, 696	0.28	.841	< 0.01		
Timing (pre/post test)	1, 696	3.98	.047	0.01		
Interaction	3, 696	2.32	0.74	0.01		
Complacency						
Experimental condition	3, 696	2.79	.040	0.01		
Timing (pre/post test)	1, 696	26.23	$< .001$	0.04		
Interaction	3, 696	0.70	.551	< 0.01		
Collective						
Experimental condition	3, 696	0.42	.738	< 0.01		
Timing (pre/post test)	1, 696	20.48	$< .001$	0.03		
Interaction	3, 696	1.35	.256	0.01		
Booster willingness						
Experimental condition	3, 696	1.05	.372	< 0.01		
Timing (pre/post test)	1, 696	6.59	.010	0.01		
Interaction	3, 696	3.63	.013	0.02		

References

Betsch, C., Schmid, P., Heinemeier, D., Korn, L., Holtmann, C., & Böhm, R. (2018). Beyond confidence: Development of a measure assessing the 5c psychological antecedents of vaccination. *PLOS ONE*, *13*, e0208601. https://doi.org/10.1371/journal.pone.0208601

Dugan, E., Trachtenberg, F., & Hall, M. A. (2005). Development of abbreviated measures to assess patient trust in a physician, a health insurer, and the medical profession. *BMC Health Services Research*, *5*, 1–7. https://doi.org/10.1186/1472-6963-5-64

Fernbach, P. M., Rogers, T., Fox, C. R., & Sloman, S. A. (2013). Political extremism is supported by an illusion of understanding. *Psychological Science*, *24*, 939–946. https://doi.org/10.1177/0956797612464058

Haun, J., Noland-Dodd, V., Varnes, J., Graham-Pole, J., Rienzo, B., & Donaldson, P. (2009). Testing the BRIEF health literacy screening tool. *Federal Practitioner*, *26*, 24–31.

Minson, J., Hagmann, D., & Luo, K. (2023). Cooling heated discourse: Conversational receptiveness boosts interpersonal evaluations and willingness to talk. https://doi.org/10.31219/osf.io/5w3dg

Primi, C., Morsanyi, K., Chiesi, F., Donati, M. A., & Hamilton, J. (2016). The development and testing of a new version of the cognitive reflection test applying item response theory (IRT). *Journal of Behavioral Decision Making*, *29*, 453–469. https://doi.org/10.1002/BDM.1883

Yeomans, M., Kantor, A., & Tingley, D. (2018). The politeness package: Detecting politeness in natural language. *R Journal*, *10*(2).

Yeomans, M., Minson, J., Collins, H., Chen, F., & Gino, F. (2020). Conversational receptiveness: Improving engagement with opposing views. *Organizational Behavior and Human Decision Processes*, *160*, 131–148. https://doi.org/10.1016/j.obhdp.2020.03.011