Computers in Human Behavior 176 (2026) 108870

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

# Countering division with friendliness: How feeling understood by a friendly AI triggers both openness and resistance

Raphael Emanuel Huber *

*University of Bern, Switzerland*

ARTICLE INFO

*Keywords:*
Feeling understood
AI chatbot
Polarization
Persuasion
Elaboration likelihood model
Misinformation
Openness
Division
Conversational receptiveness
Conspiracy theories
Cognitive warefare
Feeling heard
Counterattitudinal behavior

ABSTRACT

Fostering understanding across divides could counter societal polarization but also risks reinforcing existing views. Using AI chatbots to experimentally induce interpersonal experiences, it was tested whether feeling understood enhances openness to opposing information. Across four studies ($N = 1839$), participants engaged in 6-min chatbot conversations before reading articles that challenged their vaccination or climate change views. These conversations were experimentally varied to be neutral, deflective, friendly, corrective, or understanding-focused. Both understanding-focused and friendly conversations increased feelings of understanding, enhancing perceived credibility of opposing information and predicting counter-attitudinal behavioural intentions—mediated paths that persisted at a 60-day follow-up. While experimental mediation effects were small but consistent ($\beta = 0.02$–$0.06$), correlational relationships were robust ($\beta = 0.12$–$0.41$). The effects were strongest among anti-vaccination and climate-sceptic participants. However, despite producing the strongest initial effects on credibility and intentions, random topic friendly conversations also generated suppressed negative direct effects that became apparent at follow-up, suggesting dual systems: one that automatically responds to social cues, and another that simultaneously detects inauthenticity. Factual corrections, initially without impact, showed positive effects at follow-up. These findings illuminate feeling understood as fundamental to bridging divides—powerful enough that even artificial displays activate openness, yet suggesting potential resilience through simultaneous authenticity detection, with implications for defending against cognitive warfare.

## 1. Introduction

How can problems like climate change or a pandemic be addressed if parts of affected populations fail to recognise these issues? What purpose do conventional weapons serve when cognitive warfare occurs in people's minds, turning populations against each other? The failure of population segments to acknowledge critical global issues presents a barrier to effective societal response (Begum et al., 2024; Chan et al., 2024; Gao, 2023; Hornsey et al., 2018; Imhoff & Lamberty, 2020; Kordestani et al., 2023; Loomba et al., 2021; Nwokolo, 2025; Rutjens & Hornsey, 2024; Tyson et al., 2023; Većkalov et al., 2024).

These issues are exploited by cognitive warfare—operations targeting the human mind to manipulate how individuals think and act (Claverie, 2024; Claverie & Du Cluzel, 2022; Henschke, 2024; Pocheptsov, 2018). Such manipulation unfolds on digital platforms where misinformation and echo chambers leverage cognitive vulnerabilities to sow division (Jarynowski et al., 2023; Mahjob & Shakori, 2022; Parezanović & Proroković, 2024; Tashev et al., 2019). Polarised movements, in particular, can attract individuals with higher narcissistic and psychopathic tendencies who hijack these causes to satisfy their own ego-focused needs (Bertrams & Krispenz, 2025; Krispenz & Bertrams, 2024), potentially becoming unwitting instruments of divisive agendas. The resulting polarization can turn population segments against one another, potentially weakening societies geopolitically (U.S. Senate Select Committee on Intelligence, 2017; Zilinsky et al., 2024). This raises fundamental questions about preserving societal cohesion when warfare operates within minds themselves (Deppe & Schaal, 2024; Orinx & Struye de Swielande, 2022; Tashev et al., 2019).

### 1.1. From fact-checking to interpersonal understanding

Current research illuminates who believes misinformation and conspiracy theories and why, while revealing limitations in existing interventions. Belief in conspiracy theories correlates with social disadvantage, societal discontent, and polarised ideologies (Biddlestone et al., 2022; K. M. Douglas et al., 2017, 2023), alongside individual traits

* Fabrikstrasse 8, 3012 Bern, Switzerland
*E-mail address:* raphael.huber@unibe.ch.

https://doi.org/10.1016/j.chb.2025.108870
Received 8 July 2025; Received in revised form 31 October 2025; Accepted 15 November 2025
Available online 19 November 2025

including heightened insecurity, intuitive cognitive styles, and narcissistic tendencies (Bertlich et al., 2025; Biddlestone et al., 2022; Cosgrove & Murphy, 2023; K. M. Douglas et al., 2016, 2019, 2023; Dyrendal et al., 2021; Ecker et al., 2022; Tam & Kim, 2023). These beliefs may serve psychological functions—addressing epistemic, existential, and social needs—and potentially reflect evolved mechanisms for communicating unrepresented threats (K. M. Douglas et al., 2017; Palecek & Hampel, 2024).

Preventive interventions have employed "pre-bunking" based on inoculation theory (W. J. McGuire, 1961), exposing individuals to weakened misinformation doses to build cognitive resistance (Lewandowsky et al., 2012; Lewandowsky & Van Der Linden, 2021; Traberg et al., 2022). This approach, whether topic-specific or technique-focused, demonstrates success experimentally (Basol et al., 2020; Cook et al., 2023; Maertens et al., 2021; Roozenbeek et al., 2020, 2022; Traberg et al., 2022). While more effective than post-hoc debunking both approaches remain fundamentally reactive.

The epistemic challenge runs deeper. Establishing and communicating truth becomes problematic when conspiracy theories resist falsification and those debunking lack specific expertise (Harambam, 2021; Kirmayer, 2024a, 2024b; Pigden, 2024; Zembylas, 2023). From a critical rationalist perspective, claiming dominance over truth is inherently problematic (Popper, 1984). Some researchers therefore advocate shifting focus from direct refutation to alternative strategies that address ethical and political implications rather than narrowly epistemic concerns (Harambam, 2021; Zembylas, 2023). This epistemic humility—acknowledging that knowledge approaches rather than attains truth—may serve as fertile ground for more nuanced interpersonal approaches.

When we relinquish claims to absolute truth, we create space for genuine dialogue. Researchers exploring interpersonal strategies recommend five approaches when engaging with conspiracy believers, noting limited empirical validation for these holistic methods (K. M. Douglas et al., 2024). These include: open-minded, non-confrontational starts that increase correction credibility (Walter & Tukachinsky, 2020); affirming critical thinking while encouraging its application to believers' own sources; addressing underlying psychological needs; recognizing when to disengage; and fostering conversational receptiveness through empathy and verbal acknowledgment.

The fifth strategy—receptiveness—deserves particular attention given that division itself represents a core mechanism of cognitive warfare (Jarynowski et al., 2023; Mahjob & Shakori, 2022; Parezanović & Proroković, 2024; Tashev et al., 2019). Rather than fact-checking or preparing for misinformation, efforts could focus on directly counteracting division by fostering interpersonal understanding. This could entail listening to individuals, respecting their decisions, seeking to understand their perspectives, and acknowledging their concerns. While this approach may seem to risk amplifying extreme viewpoints—and appears counterintuitive when facing socially unaccepted views—evidence indicates that acknowledging rather than directly countering opposing opinions increases openness to different viewpoints, fosters perspective convergence, and reduces societal division.

Conversational receptiveness—specifically "receptiveness to opposing views"—is considered distinct from agreement or responsiveness, and involves acknowledging opposing views without necessarily supporting them (Minson & Chen, 2022). Research shows, teachable linguistic behaviours convey such acknowledgment, improving perceptions of reasonableness and trustworthiness (Minson & Chen, 2022; Yeomans et al., 2020), de-escalating conflict, and fostering future engagement (Itzchakov et al., 2017; Reschke et al., 2020).

### 1.2. The psychological power of feeling understood

Indeed, the subjective experience of feeling understood through such receptiveness and other means demonstrates effects across contexts. At interpersonal levels, it buffers marital conflict (Gordon & Chen, 2016), enhances intellectual humility (Itzchakov et al., 2024), and increases well-being through life satisfaction, positive affect, and fewer negative physical symptoms (Lun et al., 2008; Morelli et al., 2014; Oishi et al., 2010). At intergroup levels, it fosters trust between societal groups (Ioku & Watamura, 2022; Livingstone, Fernández Rodríguez, & Rothers, 2020; Livingstone, Windeatt, et al., 2020), promotes cross-group contact intentions (Itzchakov & Reis, 2021; Roos et al., 2023), and facilitates cooperation across cultural divides (Bruneau & Saxe, 2012; Ioku & Watamura, 2022, 2025; Livingstone, 2023; Oishi et al., 2010).

These benefits manifest neurally. Feeling understood activates reward-associated regions like the ventral striatum, while being misunderstood engages the anterior insula and networks processing social pain (Morelli et al., 2014; Seehausen et al., 2014). Such experiences satisfy fundamental needs for recognition and belonging, fostering psychological safety that reduces defensiveness and reactance when discussing contentious topics (Gordon & Chen, 2016; Itzchakov & DeMarree, 2022; Minson & Chen, 2022).

However, "feeling understood" is lacking uniform definition and measures. Though it is often assessed through relationship-focused scales (Gordon & Chen, 2016; Reis et al., 2017), intergroup perception measures (Livingstone, Fernández Rodríguez, & Rothers, 2020), or cognitive appraisals of conversational quality (Morelli et al., 2014; Roos et al., 2023; Seehausen et al., 2014; Yin et al., 2024). The acute, primarily affective dimension of feeling understood—as a distinct emotional state immediately following specific exchanges—remains underexplored.

Similarly varied are studies measuring how feeling understood relates to openness to opposing information. They typically assess behavioural intentions to engage with opposing views (Itzchakov & Reis, 2021; Livingstone, Fernández Rodríguez, & Rothers, 2020), willingness to affiliate with those holding different perspectives (Minson et al., 2020; Yeomans et al., 2020), or shifts in attitude structure and certainty (Itzchakov et al., 2017, 2024). Few operationalise openness as direct credibility evaluation of specific opposing information immediately post-interaction, nor trace this to subsequent behavioural choices.

### 1.3. Integrating feelings of understanding into the Elaboration Likelihood Model

Furthermore, despite its relevance to persuasion and positive affect (Griskevicius et al., 2010; Petty & Briñol, 2015; Petty et al., 1993; Schwarz & Bless, 1991; Worth & Mackie, 1987), the Elaboration Likelihood Model (Petty & Cacioppo, 1986), a cornerstone of persuasion research, has not been the predominant framework for examining how feelings of understanding influence attitude change. The ELM's potential for explaining how psychological safety fosters unbiased elaboration was noted, while simultaneously critiquing its traditional focus on unidirectional rather than interpersonal communication (Itzchakov & DeMarree, 2022). However, the ELM can accommodate interpersonal dynamics when information processing is understood as continuously adapting based on reciprocal experiences of validation within the interaction.

Building on how positive affect influences elaboration (Petty & Briñol, 2015; Petty et al., 1993), four potential mechanisms can be outlined—partially challenging findings that positive mood promotes peripheral processing (Schwarz & Bless, 1991; Worth & Mackie, 1987)—through which feeling understood might facilitate more favorable processing of opposing information (Fig. 1).

First, feeling understood could enhance motivation for effortful central route processing. The empirically demonstrated psychological safety and interpersonal connection from feeling understood (Gordon & Chen, 2016; Itzchakov & DeMarree, 2022) transforms potentially adversarial exchanges into constructive dialogue, increasing willingness to carefully consider challenging information.

Second, opposing information could trigger negative affect that depletes cognitive resources. Feeling understood activates reward regions

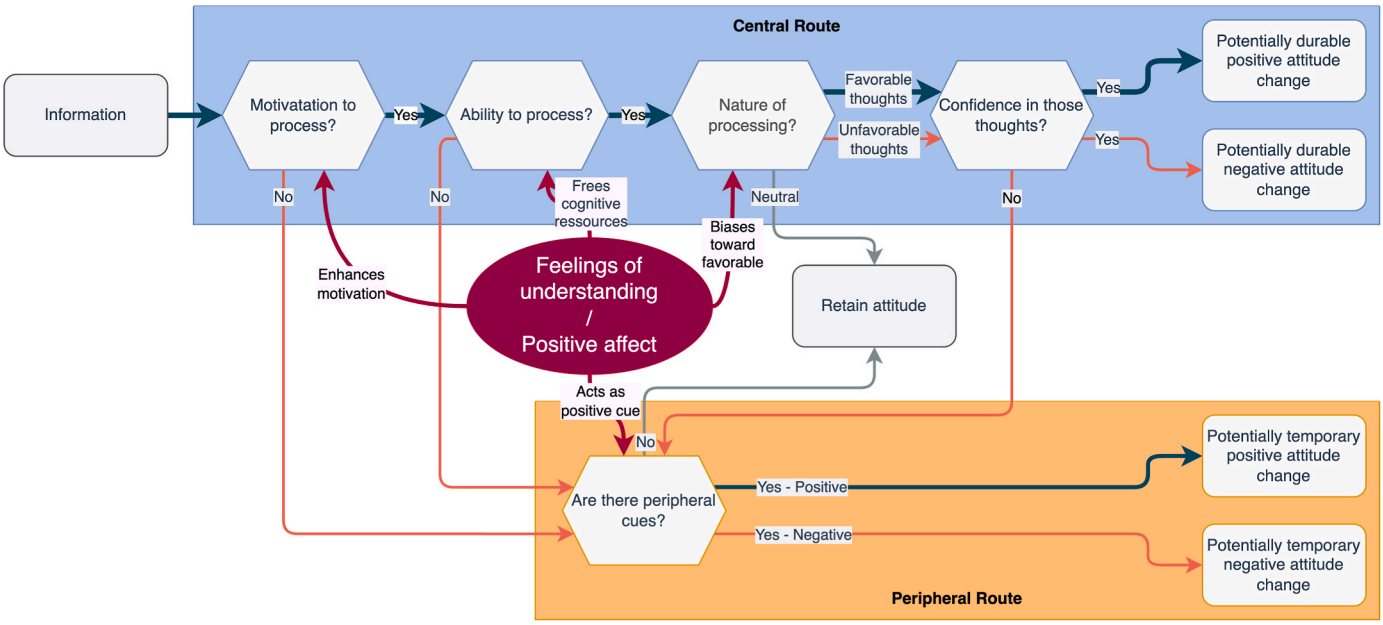


**Fig. 1.** Extended elaboration likelihood model incorporating feelings of understanding.
*Note.* Thicker green arrows indicate positive/favorable processing paths, thin red arrows indicate negative/unfavourable paths, grey arrows indicate neutral outcomes, and purple arrows represent the theoretical contributions of feelings of understanding to processing outcomes.

(Morelli et al., 2014), potentially mitigating these responses and freeing cognitive capacity for systematic processing.

Third, feeling understood could influence the nature of elaboration during central route processing. Rather than defensive counter-arguing, enhanced source trustworthiness from feeling understood can steer cognitive processing toward more thorough, less biased assessment—particularly when elaboration likelihood is high.

Fourth, under low elaboration conditions, feeling understood may act as a peripheral cue. The positive affective state may be misattributed to the message itself, while enhanced source likability encourages acceptance through heuristic processing ("this person understands me, so their message is likely valid").

### 1.4. The present research

The present research program addressed these gaps by investigating how experimentally induced feelings of understanding influence engagement with opposing information. Openness was operationalised through immediate credibility ratings and downstream counter-attitudinal behavioural intentions and define feeling understood as the positive affective state experienced when perceiving conversational acknowledgment of one's viewpoint.

It was hypothesised that feeling understood influences perceived credibility of opposing information, which in turn mediates effects on behavioural intentions. Four studies progressively explored this phenomenon across vaccination (Pilot, Study 1) and climate change (Studies 2–3) contexts. The general procedure involved: measuring baseline attitudes; randomly assigning conversational experiences; assessing feelings of understanding (Studies 1–3); presenting counter-attitudinal news articles; measuring perceived credibility; and assessing behavioural intentions (vaccinate, vote for pro-climate politician) through fictional scenarios (Fig. 2).

Given potential complications of human-led experimental conversations, Large Language Model chatbots were employed based on the Computers Are Social Actors paradigm (Reeves & Nass, 1996), which posits that individuals unconsciously apply social norms to computer interactions.

The Pilot study validated this methodology, comparing no conversation with an understanding chatbot that explored participants' vaccination opinions while remaining non-confrontational and unconditionally acknowledging, aiming to induce feelings of understanding related to a topic to positively influence central route processing. Study 1 directly contrasted this understanding approach with deflecting responses on unrelated topics, introducing the Feelings of Understanding/Misunderstanding Scale (Cahn & Shulman, 1984) to capture immediate affective experiences. Study 2 extended findings to climate sceptics, comparing four conditions: neutral conversation on random topics, corrective conversation countering climate misconceptions with scientific arguments, understanding conversation exploring their climate views non-confrontationally, and friendly conversation on positive unrelated topics to induce topic unrelated positive emotion potentially triggering peripheral processing. This design differentiated feeling understood about one's views from factual correction or general positivity. Study 3 provided 60-day longitudinal follow-up.

Across studies, LLM technology's potential was explored for qualitative conversation evaluation and data quality assessment.

This research program extends existing work on interpersonal understanding by examining how feeling understood influences information processing—without amplifying previous viewpoints. By providing empirical evidence for these mechanisms, the study sheds light on how

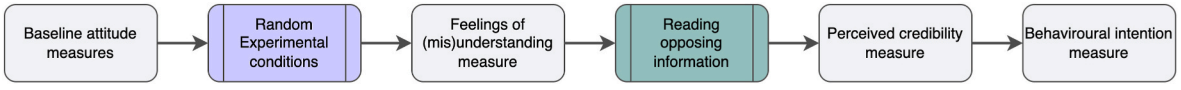


**Fig. 2.** Experimental procedure across studies.
*Note.* The Feelings of (Mis)Understanding Scale was only administered in Studies 1–3. Experimental conditions: Pilot (No Chat vs. Understanding Chat), Study 1 (Deflective vs. Understanding Chat), Study 2 (Neutral vs. Corrective vs. Understanding vs. Friendly Chat). Study 3 was a 60-day follow-up without manipulation re-measuring attitudes and intention.

interpersonal strategies can preserve democratic discourse and societal cohesion without claiming epistemic dominance. This addresses division at its core rather than its symptoms, offering a path beyond perpetual fact-checking battles.

## 2. Methods

### 2.1. Common methods across studies

**Ethics and Overview:** Ethical clearance was obtained from the Research Ethics Committee at the University of __ (2022-07-0004) [Anonymized for Peer Review]. All participants provided informed consent. Data, analysis code, and software version details are available at https://osf.io/tn4cq/and Supplementary Methods S2.1.2. Throughout, 'SM' denotes Supplementary Materials, 'SMt' Supplementary Methods, and 'SR' Supplementary Results. Single references starting with 3 indicate parallel content across studies (e.g., SM-S3.2, S4.2, S5.2).

**Participants and Recruitment:** Participants aged 18+ from the United States were recruited via CloudResearch Connect (Hartman et al., 2023) for all studies and Prime Panels (Chandler et al., 2019) additionally for Study 2. Compensation ranged from $2.00–2.50 for 14–15 min surveys. Sample sizes determined through Monte Carlo simulations (Kline, 2023; Muthén & Muthén, 2002) for planned structural equation models, supplemented by guidelines: 10:1 participant-to-parameter ratio (Bentler & Chou, 1987; Brown, 2015; Hoyle, 2023; Kline, 2023; Muthén & Muthén, 2002; Whittaker & Schumacker, 2022). Power analyses targeted 80 % power (α = .05, two-tailed) for detecting medium effect sizes based on prior research (Jolley & Douglas, 2014, 2017). Final samples after quality exclusions: Pilot ($n$ = 239), Study 1 ($n$ = 470), Study 2 ($n$ = 752), Study 3 ($n$ = 378). Details per Study: SMt-S3.2.2.

**General Procedure:** Studies 1–3 followed consistent structure: baseline attitude measurement, 6-min chatbot conversation (experimental manipulation), post-manipulation assessment including feelings of understanding (Study 1–3), presentation of attitude-challenging article, credibility assessment, and behavioural intention measurement. Attention checks were embedded throughout (SMt-S2.2). Chatbot interactions used OpenAI's API (GPT-4 for Pilot/Study 1; GPT-4o for Study 2) embedded via iFrame. Conversations automatically terminated after 6 min. All participants received welcoming messages emphasizing genuine interest and non-judgment to reduce social desirability bias. Chatbot introductions were minimal across all conditions ("converse with the Chat-Bot below … Read your answers later"), except Understanding Chat which received framing emphasizing genuine interest, appreciation for viewpoints, and the chatbot's training on real interactions. This differential framing aimed to amplify the Understanding Chat's subsequent empathetic responses. All participants received comprehensive appropriate debriefing. Full procedure: SM-S3.2 and SMt-S3.3. Given documented quality issues with online participant platforms (B. D. Douglas et al., 2023; Peer et al., 2021; Webb & Tangney, 2024; Zhang & Gearhart, 2020), a multi-layered quality assurance system was implemented across all studies. An extensive rationale for and detailed criteria are provided in SMt S2.2.

**Core Measures:** *Feelings of Understanding/Misunderstanding (FUM) Scale (Studies 1–3)*: Measures affective experience of feeling understood versus misunderstood (Cahn & Shulman, 1984) through ratings of associated feelings (satisfaction, comfort versus annoyance, sadness) on 5-point scales. Chosen deliberately because critiqued for measuring feelings rather than a cognitive appraisal (Grice, 1997; Schrodt, 2003; Schrodt & Finn, 2011). Scores calculated as understanding minus misunderstanding ratings (α = .92- 0.94).

*Perceived Article Credibility:* Five items adapted from Flanagin & Metzger (Flanagin & Metzger, 2000) assessing accuracy, trustworthiness, completeness, bias, and overall credibility on 7-point scales (α = .88-0.93).

*Topic-Specific Measures:* The Vaccination Status Identification (Henkel et al., 2022) (VSI) scale Studies 1 and 2; Vaccination Attitudes Examination (Martin & Petrie, 2017) (VAX) for Study 1; Climate Change Scepticism Questionnaire (CCSQ) for Studies 2–3 (De Graaf et al., 2023), measuring on 6/7-point scales (α = .79- 0.91).

### 2.2. Complete measure listings, descriptions, and characteristics appear in SM-S3.2, SMt-2.10 and SR-S3.3ff

**Chatbot Implementation:** The Understanding Chat condition instructed chatbots to enquire about participants' topic (vaccination, climate change) opinions then respond with unconditional friendliness and understanding. Core instructions included using phrases like "I understand," asking clarifying questions, and finding common ground while explicitly prohibiting factual corrections or opposition. This required informing GPT-4 of the research context to override default correction behaviours (complete prompts in SM-S3.2.4). See Table 1 for a listing of all experimental conditions across studies.

**Statistical Analysis:** Analyses progressed from descriptive statistics, correlation tables, through regression to structural equation modelling. Group comparisons procedures SMt-2.3. Regressions and Structural Equation Modelling (SEM) progressed from base models to extended models with demographic control (sex, age, education, race, SMt-2.5) and exploratory control variables (SMt-2.11). The full set of control variables was included in regressions, which subsequently informed their selective inclusion in the SEM. Ordinal logistic regression handled Likert outcomes. Assumptions were checked and where violated appropriate methods were employed (i.e. scale parameters allowed to vary (Agresti, 2010; Tutz, 2022), MM-estimation, robustbase (Maechler et al., 2024), KS 2014 (Koller & Stahel, 2017) to reduce local breakdown issues). SEM using lavaan (Rosseel et al., 2024) tested the hypothesised mediation pathway derived directly from our theoretical extension of the Elaboration Likelihood Model (ELM), as detailed in Fig. 1: experimental condition → feelings of understanding → article credibility → behavioural intentions. Study 1 used WLSMV estimation appropriate for ordinal outcomes (Brown, 2015; Sass et al., 2014; Whittaker &

**Table 1**
Overview and description of experimental conditions across all studies.

| Condition | Topic | Brief Description | Studies ($n$) |
|---|---|---|---|
| No Chat | N/A | Participants waited for 1 min; served as a no-interaction baseline. | Pilot (129) |
| Deflective | Unrelated | Mimicked a non-validating partner by asking superficial questions and ignoring participant responses. | Study 1 (253) |
| Neutral | Unrelated | Asked generic, indifferent-toned questions on random topics, establishing a non-emotional baseline. | Study 2 (207) |
| Understanding | On-Topic | Explored and validated participant's specific views (on vaccination or climate change) non-confrontationally. | Pilot (110) Study 1 (217) Study 2 (165) |
| Corrective | On-Topic | Respectfully countered participant's misconceptions with scientific facts and arguments. | Study 2 (181) |
| Friendly | Unrelated | Engaged in a warm, positive conversation on participant-chosen topics, validating the *person* rather than their views. | Study 2 (199) |
| Condition | Topic | Brief Description | Studies (n) |

*Note.* $N$ = number of participants in the final sample for each condition. "On-Topic" refers to conversations about the core study topic (vaccination in Pilot/Study 1; climate change in Study 2); "Unrelated" refers to conversations on different topics. Total $N$ = 1839. Full chatbot prompts, fine-tuning data, and detailed implementation logic for each condition are available in the Supplementary Materials and Methods.

Schumacker, 2022). Potential WLSMV fit indices inflation (Kline, 2023; C.-H. Li, 2016, Rhemtulla et al., 2012; Savalei, 2021) was addressed through sensitivity analyses using MLR. Studies 2–3 utilised MLR (Maximum Likelihood Robust) estimator appropriate for multi-item outcomes and non-normal distributions (Brown, 2015; Hoyle, 2023; Sass et al., 2014); Model fit assessed via CFI/TLI (>0.90), RMSEA (<0.08), and SRMR (<0.08) (Hu & Bentler, 1999; Kline, 2023). Commentary on treating Likert scales as continuous: SMt-S2.4. Baseline attitude scores (VAX for vaccination, CCS for climate) were controlled in all models as they determined participant inclusion and strongly predicted outcomes. Controlling for these pre-existing opinions isolated experimental effects from baseline differences, preventing omitted variable bias and improving treatment effect precision (Hayes & Little, 2022; Kahan & Morris, 2012).

**Large Language Model Analysis:** LLMs systematically analysed chat conversations for participant ID extraction and engagement assessment. Implementation evolved from basic GPT-4 (Pilot) to multi-model validation including fine-tuned models (Study 2). Human validation served as gold standard throughout, with minimal exclusions resulting from chat engagement assessment. For the Pilot and Study 1 all conversations were evaluated by one human reviewer. In Study 2 a hybrid approach was tested: All conversations flagged by the LLM for poor engagement were manually reviewed, additionally random 10 % of the remaining conversations (rated unanimously positive by multiple LLM reviewers) were human reviewed. Any conversation confirmed as non-serious or meaningless by the human rater was excluded. This hybrid approach ensured data quality while managing the large volume of conversational data. Exclusion due to low quality chat was 0.7 %–1.5 % (See SMr, 3.3.2, 4.3.3, 5.3.3 for extensive exclusion analysis). Furthermore, LLM analysis methods and additional explorations are comprehensively reported in SMt-2.12, SM-S3.3. Interrater reliability between human review and different LLMs are reported in SMr-3.6, 4.6, 5.6.

**Deviations from Preregistrations:** The analysis was guided by the preregistrations, with several refinements implemented to ensure the most rigorous test of the hypotheses. These refinements included:

Statistical Modelling: To best fit the observed data, robust and ordinal regression methods were employed instead of standard GLMs. Baseline attitudes were incorporated as covariates to isolate experimental effects more precisely, a necessary step given their strong predictive power.

Data Quality Assurance: Preregistered qualitative data checks were supplemented with quantitative criteria (e.g., for chat engagement and outlier detection) to ensure data validity.

Sampling: Practical sampling strategies were adapted as needed to meet recruitment targets.

A complete description and justification for all refinements across each study is documented in the Supplementary Materials (SMt-3.6).

### 2.3. Pilot study

**Participants:** From 317 participants pre-selected for COVID-19 non-vaccination status recruited via Connect (May 2024), 300 completed the study. Following multi-layered quality control, 61 participants (20.3 %) were excluded—primarily for reading 300-word articles under 30 s ($n$ = 37) or chat-related criteria ($n$ = 15). Excluded participants were more likely male, younger, with faster reading times (all P values < 0.05). Final sample: $N$ = 239 (mean age = 41.3 years, $SD$ = 12.5; 63.2 % female). See SM-S3.2 and SR-S3.2ff for complete flow and exclusion analyses.

**Procedure:** Participants were informed about real-time chat, requiring comfort with typing longer text. After baseline scales (VSI, VCB, PDV) random assignment placed participants in either No Chat (1-min wait) or Understanding Chat conditions. They were told they would communicate with researchers but not informed of AI involvement—a deception addressed only during debriefing. Real-time scoring of combined vaccine conspiracy and danger beliefs (>76 = anti-vaccination) determined article assignment: pro-vaccination participants read conspiracy content (Jolley & Douglas, 2014) claiming vaccines cause autism and pharmaceutical profit motives, while anti-vaccination participants read debunking content with scientific evidence. The vaccination scenario asked participants to assume parenthood of 8-month-old "Sophie" facing a decision about vaccinating against "dysomeria"—a fictional disease causing fever, vomiting, and potential severe outcomes (Betsch & Sachse, 2013; Jolley & Douglas, 2014). Pro-vaccination participants who read conspiracy content received corrective information during debriefing.

**Materials** (SMt-2.10)**:** *Vaccine Conspiracy Belief Scale (VCB):* Eight items measuring anti-vaccine conspiracy endorsement (Jolley & Douglas, 2014, 2017) on 7-point scales (α = .96/ω = 0.96). *Perceived Dangers of Vaccines Scale (PDV):* Eight items assessing vaccine risk beliefs (Betsch & Sachse, 2013; Jolley & Douglas, 2014, 2017) on 7-point scales (α = .88/ω = 0.88). *Chat Satisfaction:* Three custom items (enjoyment, feeling understood, respectful treatment) for Understanding Chat participants only (α = .77/ω = 0.80). *Perceived article credibility* (Flanagin & Metzger, 2000) (α = .93/ω = 0.94). *Vaccination Intention:* Single item assessing intention to vaccinate (1 = definitely not, 7 = definitely vaccinate).

**Analysis:** Analyses used ordinal logistic regression for vaccination intention (categorical outcome) and linear models for credibility. Heteroscedasticity in vaccination models, conditional on age and vaccination opinion interactions, required heteroscedastic cumulative link models. Given modest condition × group interaction evidence (p = .041), Bayesian analysis provided additional inference. Complete procedures are in SMt-S3.4.

Exploratory analyses examined chat satisfaction and AI recognition as credibility predictors. Chat satisfaction showed initial significance (β = 0.21, $p$ = .018) that became non-significant with controls (β = 0.15, $p$ = .110). VSI showed no explanatory power beyond VCB and PDV, though surveying vaccination status identification might have fortified positions—prompting its retention in Study 1.

### 2.4. Study 1

**Rationale:** Seven modifications addressed pilot limitations to enhance methodological rigor. (SMt-3.7 for detailed rationale).

*AI Disclosure:* Participants were informed of AI interaction upfront, addressing negative effects when pilot participants discovered deception. This prevents expectancy violations (Burgoon, 1993), follows emerging ethical guidelines (Bloch-Atefi, 2025; J. McGuire et al., 2023), and reduces confounding from trust-breach reactions that could influence information processing pathways (Petty & Briñol, 2015; Petty & Cacioppo, 1986).

*Active Control:* "Deflective Chat" replaced no-interaction control, as non-responsive chatbots produce lower perceived understanding than empathetic responses (Rheu et al., 2024). Contrasting understanding with active deflection maximized variance in subjective experiences, enabling direct mechanism testing while controlling for interaction effects.

*FUM Scale:* Introduced Feelings of Understanding/Misunderstanding Scale to empirically measure the subjective experience previously only assumed, strengthening internal validity (Andrade, 2018; Döring & Bortz, 2016).

Unified Measurement: A single vaccination attitude scale replaced two pilot scales, enhancing measurement quality (Andrade, 2018; Döring & Bortz, 2016).

*Contemporary nuanced Materials:* Articles shifted from polarised vaccination arguments to COVID-19 mRNA vaccines (55 % of pilot conversations mentioned COVID-19). It replaced conspiracy accusations with doubt-fostering arguments about development speed; pro-vaccine content acknowledged concerns while presenting statistical evidence. This was implemented to avoid defensive reactions from extreme framing that may prevent central route processing (Velez & Liu, 2024;

Wood & Porter, 2016).

*Aligned Scenario:* The vaccination scenario was modified to specify mRNA vaccines, matching updated reading materials.

*Neutral Exclusion:* Neutral-stance participants were excluded to ensure clear baseline opinions, creating homogeneous groups with reduced variability and increased statistical power (Heidel, 2016; Jager et al., 2017).

**Participants:** From 1089 participants recruited via Connect (September 2024), 396 (35.0 %) were screened for neutral vaccination opinions and 36 (5.1 %) dropped out. From 657 completers, 187 (28.5 %) were excluded—primarily for reading 270-word articles under 25 s ($n = 99$) or chat criteria ($n = 56$). Excluded participants were younger, Hispanic/Latino, less educated, lower income, rating articles as more credible (all $p$ values $< 0.05$). Final sample: $N = 470$ (mean age = 42.6 years, $SD = 12.3$; 59.6 % female). See SR-S4.2, SM-S4.2.

**Procedure:** Participants were explicitly informed about chatbot interaction. After VSI (Henkel et al., 2022) and VAX scales (Martin & Petrie, 2017), neutral-scoring participants (VAX 31–53) were screened out. Random assignment: Understanding Chat (identical to Pilot) or Deflective Chat (ask superficial non-vaccination questions, maintain robotic tone, disregard responses, statements with contrastive conjunctions. See SM-S4.2.4 for prompts). Post-chat, participants completed FUM scale (Cahn & Shulman, 1984), read custom COVID-19-specific moderately opposing articles, assessed credibility (Flanagin & Metzger, 2000), and responded to modified scenario specifying "mRNA vaccine." See SM-S4.2 and SMt-S4.3.

**Materials** (SMt-2.10): *(VAX) (*Martin & Petrie, 2017): Replacing the Pilot's two scales VCB and PDV, measured vaccination attitudes ($\alpha = .98/\omega = 0.98$). *(FUM* (Cahn & Shulman, 1984): ($\alpha = .92/\omega = 0.92$). Perceived article credibility (Flanagin & Metzger, 2000): ($\alpha = .93/\omega = 0.94$).

**Analysis:** Ordinal logistic regression for vaccination intention was employed as outlined in SMt-S4.4 and failed in the Pro-Vaccination group due to proportional odds violations and sparse lower categories ($n < 6$ for ratings 1–3). Alternative approaches (Bayes) failed to converge, requiring cautious interpretation. Most linear models (on FUM, and credibility) needed robust estimation due to assumption violations.

Exploratory segmented regression identified breakpoints where chat turns-FUM relationships changed: 15 turns (Pro-Vaccination) and 11 turns (Anti-Vaccination). Article reading duration was added as a control based on correlations. VSI again showed no explanatory power.

Multi-group SEM tested COND→ FUM→ CRED→ VACINT with VAX as covariate. Feelings of (mis)understanding was specified as second-order factor indicated by first-order latent factors feelings of understanding (eight items) and feelings of misunderstanding (eight items), reflecting the conceptualization of the scale as the difference between these constructs. After removing two poor-loading misunderstanding items (<0.45), partial metric invariance was established, confirming comparability of coefficients between the two groups (pro/anti-vaccination). The model with controls showed improved fit and was selected. Sensitivity analysis with MLR and path model supported the WLSMV model findings (complete description SMt-S4.5).

### 2.5. Study 2

**Rationale:** Nine modifications enhanced methodological rigor and theoretical scope based on pilot and Study 1 insights (SMt-4.7 for detailed rationale).

*Topic Shift:* The topic was changed from vaccination to climate change to decouple findings from COVID-19-specific confounds (media saturation, political alignments). This strengthens confidence that effects relate to feeling understood rather than topic artifacts (Andrade, 2018; Döring & Bortz, 2016).

*Sceptic-Only Recruitment:* Exclusively climate sceptics were recruited after Pro-vaccination participants showed minimal effects, enabling targeted mechanism examination.

*Nuanced Materials:* The pro-consensus climate article acknowledged natural climate variations before presenting evidence, avoiding defensive reactions from extreme framing (Velez & Liu, 2024). It incorporated value-framing connecting climate action to conservative principles (self-reliance, energy independence) (Feinberg & Willer, 2013; Lakoff, 2002; Wolsko et al., 2016).

*Chatbot Authorship:* The article was attributed to the chat partner per CASA paradigm (Gambino et al., 2020; Nass & Moon, 2000; Reeves & Nass, 1996), enabling direct testing of how feeling understood by a source influences processing their message (Petty & Briñol, 2015; Petty & Cacioppo, 1986).

*Multi-Item Voting:* The single-item intention was replaced with a custom four items voting intention, and facilitating more robust statistical analyses (DeVellis, 2017). The candidate to be voted for was profiled as the chat partner, from the participant's party but centrist/pro-climate, creating tension between party loyalty and issue scepticism.

*New Conditions:* Deflective Chat was replaced with three conditions to isolate mechanisms. Neutral Chat asked irrelevant questions without eliciting emotions. Corrective Chat politely addressed climate misconceptions with scientific evidence, testing whether factual correction works (Costello et al., 2024), though effectiveness was questioned (Lisker et al., 2025). Friendly Chat engaged participants in warm, climate-unrelated conversation to test whether general positive affect differs from topic-specific understanding—distinguishing superficial friendliness from the deeper engagement of Understanding Chat. Understanding Chat was hypothesised to provide topic-specific validation, potentially engaging central-route processing, whereas 'Friendly Chat' was designed to induce general positive affect, hypothesised to act as a peripheral cue (Petty & Cacioppo, 1986).

*Fine-Tuning:* LLMs were trained on example conversations for improved coherence versus single shot prompting.

*Anthropomorphism Enhancement:* Two post-chat items instructed participants to imagine speaking with a real person and rate friendship/confiding potential. These served as methodological tools amplifying social processing tendencies per CASA principles (Fox & Gambino, 2021; Gambino et al., 2020; Heyselaar, 2023; Nass & Moon, 2000)—strengthening both automatic and mindful anthropomorphic responses to technology (Gu et al., 2024; Q. Li et al., 2023; Rao Hill & Troshani, 2024). This priming potentially enhanced ecological validity of subsequent article responses.

*Comprehension Checks:* Six content-based items ensured article processing, addressing limitations of simple attention checks (Guerreiro et al., 2022; Muszyński, 2023; Shamon & Berning, 2020).

**Participants:** From 4647 participants recruited via Connect and Prime Panels (pre-targeting climate-uncertain users, then Republicans as scepticism proxy (Ballew et al., 2019; De Graaf et al., 2023; Hornsey et al., 2018; McCright et al., 2016; Tyson et al., 2023). December 2024–January 2025), 2430 (52.3 %) were screened as non-sceptics. Of 2217 eligible participants, 884 (39.9 %) dropped out—younger, male, Hispanic/Latino, unemployed, less educated, Prime Panels recruits (all P values $< 0.05$), no condition differences. From 1333 completers, 581 (43.6 %) were excluded—primarily for article comprehension failure ($n = 254$) and insufficient chat engagement ($n = 250$). Prime Panels showed higher dropout/exclusion than Connect (28.5 % vs 3 %; 61.8 % vs 27.0 %). Excluded participants were older, male, less educated, unemployed, lower FUM scores (all $p$ values $< 0.05$); Corrective Chat showed highest exclusion. Final sample: $N = 752$ climate sceptics (mean age = 53.2 years, $SD = 16.2$; 45.4 % female). See SMt-5.2, SM-S5.2, SR S5.2–5.3 for details on methods, flow and results.

**Procedure:** After CCSQ (De Graaf et al., 2023), climate sceptics (scores >48) were randomly assigned: (1) Neutral Chat—generic questions, indifferent tone; (2) Corrective Chat—factual climate information addressing misconceptions respectfully; (3) Understanding Chat—acknowledging climate views without contradiction; (4) Friendly

Chat—warm interaction on participant-chosen topics. All except Neutral used fine-tuned GPT-4o models to attempt improved conversational flow (fine-tune details: SM-S5.2.4).

Post-chat, participants completed anthropomorphism items ("imagine talking to real person"; rate befriend/confide potential) as methodological priming before FUM scale, not used for analysis. Crucially, participants read a climate consensus article attributed to the chat partner, using value-framing (Feinberg & Willer, 2013; Lakoff, 2002; Wolsko et al., 2016) (community, self-reliance). After credibility assessment and six comprehension questions, participants rated their voting intention for the chat partner as political candidate (participant's party, centrist, pro climate implicit through having authored the article) across four offices. See SM-S5.2, SMt-S5.3.

**Materials:** *CCSQ* (De Graaf et al., 2023): Twelve items measuring climate scepticism dimensions (α = .79/ω = 0.82). *(FUM) (*Cahn & Shulman, 1984) (α = .94/ω = 0.94). Perceived article credibility (Flanagin & Metzger, 2000) (α = .89/ω = 0.91). *Voting intention:* A custom scale comprising four items assessing likelihood to vote for the chat partner as candidate across political offices (city council, state governor, US senate, president; α = .98/ω = 0.98). *Article comprehension:* Six true/false items verifying understanding were used to assess meaningful participation.

**Analysis:** Regression analyses revealed assumption violations requiring robust MM-estimation across all models. Exploratory analysis confirmed the chat turns-FUM breakpoint from Study 1, which was then incorporated as a binary predictor in models with controls (SMt-S5.4).

The SEM tested COND→ FUM→ CRED→ VOTEINT with CCSQ as covariate employed MLR estimation due to multi-item outcomes and non-normal distributions (Brown, 2015; Hoyle, 2023; Sass et al., 2014). WLSMV convergence issues from extreme multicollinearity—the four offices correlated too highly—confirmed this choice. MLR provided stable solutions without modifications with FIML handling missing data ($n$ = 739–746). The model with controls showed improved fit and was selected. Sensitivity analysis with path model supported the findings (complete description SMt-S5.5).

### 2.6. Study 3

**Rationale:** Study 2's Understanding and Friendly Chat conditions increased pro-climate voting intentions, prompting 60-day follow-up to assess effect persistence (SMt-5.7 for detailed rationale).

*Differential Predictions:* Understanding Chat's small effects suggested potential non-detectability after 60 days, though persistence would support theoretical claims of central route processing producing durable change (Petty et al., 1995). Conversely, Friendly Chat's stronger immediate effects—likely from transient positive affect as peripheral cues—were expected to disappear once emotions subsided.

*Source Independence:* Unlike Study 2's integrated design (chatbot→article→voting for same source), Study 3 presented the political candidate without prior study reference. This tested whether attitude changes persisted absent immediate source cues. Per ELM, central route changes should endure independently, while peripheral route effects require source presence. This design assessed whether climate attitude shifts influenced behavioural intentions without apparent connection to the original chatbot interaction.

**Participants:** All 525 preliminary valid Connect participants from Study 2 were invited for 60-day follow-up (January–February 2025); Prime Panels participants could not be re-contacted. Of 752 eligible Study 2 participants, 378 (50.3 %) completed follow-up. Non-returning participants were more likely employed, from income extremes, older, more climate-sceptical, and from Corrective Chat condition (all P values $< 0.05$). From 447 accessing the survey, 43 were retrospectively excluded based on finalised Study 2 criteria. Final sample: N = 378 (mean age = 47.6 years, $SD$ = 13.7; 51.6 % female; 71.6 % retention among Connect participants). See SMt-S6.2, SR-S6.2ff.

**Procedure:** After consent and re-completing CCSQ (De Graaf et al., 2023), participants responded to a custom voting scenario. As Study 2, but the candidate was introduced independently without chatbot/article connection featuring a politician from participant's party (unnamed) with centrist views advocating climate action through conservative-resonant framing (market solutions, energy independence). No experimental manipulation, allowing assessment of Study 2 effect persistence.

**Materials:** *CCSQ* (De Graaf et al., 2023): Re-administered to measure attitude change (α = .91/ω = 0.92). *Voting intention* (α = .98/ω = 0.98).

**Analysis:** All regression models showed assumption violations requiring robust MM-estimation (SMt-S6.3). SEM tested: Study 2 COND→ FUM→ CRED→ CCSQ-t2→ VOTEINT-t2, with baseline CCSQ as covariate. Due to extreme correlation between voting items 2 and 3 (r = 0.97), increased model complexity with reduced sample size, these were removed, retaining items 1 and 4 (voting for city council, president). MLR estimation with FIML handled missing data ($n$ = 374–376). The control model selected based on improved fit. Path analysis confirmed results, though model fits were poor (SMt-S6.4, SR-6.4).

## 3. Results

Across four studies, it was tested whether AI chatbot conversations could enhance openness to opposing views. The Pilot study ($N$ = 239) established proof-of-concept with vaccination attitudes, revealing differential effects by stance. Study 1 ($N$ = 470) introduced the Feelings of Understanding/Misunderstanding (FUM) scale and confirmed the mediation pathway. Study 2 ($N$ = 752) compared four conversational strategies with climate sceptics, revealing a hierarchy of effectiveness. Study 3 ($N$ = 378) examined 60-day persistence of effects.

Despite differential attrition across studies (ranging from 20.3 % to 50.3 %) randomization remained successful in each study with no baseline differences in baseline attitude measures or demographics (Supplementary Results S3.3.4, S4.3.5, S5.3.4, 6.3.2). A demographic pattern would replicate throughout the program, women and Black/African American participants rated opposing articles as more credible. Other demographic patterns varied (See Methods and Supplementary Results S3.4, 4.4, 5.4, 6.3 for detailed attrition results and analysis). All p-values are reported as two-tailed despite directional hypotheses in Studies 1–3, maintaining conservative statistical practices.

### 3.1. Pilot

The pilot study ($N$ = 239 after 20.3 % exclusions for data quality) provided initial evidence that AI chatbot conversations acknowledging participants' vaccination views could influence subsequent vaccination intentions, particularly among anti-vaccination participants. Participants were randomly assigned to either no chat control ($n$ = 129) or Understanding Chat ($n$ = 110), where they were told they would converse with another person but actually interacted with an AI chatbot.

Without direct measurement of feelings of understanding (FUM scale not yet implemented), interaction satisfaction served as a proxy. Understanding Chat participants reported high satisfaction (mean 16.2 of 21), suggesting positive interactional experiences. The primary analysis revealed important group differences in how the manipulation affected vaccination intentions.

For anti-vaccination participants reading pro-vaccination content, Understanding Chat increased vaccination intentions (OR = 5.15, $p$ = .041), and perceived credibility strongly predicted vaccination intentions in this group as well (OR = 4.70, $p < .001$)—each standard deviation increase nearly quadrupled the odds of higher vaccination intention. Pro-vaccination participants showed different dynamics: Understanding Chat did not significantly affect their vaccination intentions ($p$ = .091), and neither did article credibility ($p$ = .290).

A crucial discovery emerged from exploratory analyses: participants who indicated having realised they were conversing with AI (despite being told it was another person) rated articles as significantly less

credible (β = −0.79, $p$ = .027), while higher chat satisfaction predicted increased credibility (β = 0.21, $p$ = .018). This finding—that perceived deception undermined effects while authenticity was critical—informed all subsequent study designs, leading us to explicitly disclose the AI nature of conversations.

### 3.2. Study 1

Study 1 ($N$ = 470) replicated and extended the pilot findings with two key improvements: implementing the Feelings of Understanding/Misunderstanding (FUM) scale and explicitly informing participants they would chat with AI. Participants with extreme vaccination views (neutrals were screened out) were randomly assigned to Deflective Chat ($n$ = 253; deflecting to unrelated topics) or Understanding Chat ($n$ = 217; exploring and acknowledging vaccination views). The distinct pro-vaccination ($n$ = 284) and anti-vaccination ($n$ = 186) groups differed systematically—pro-vaccination participants were more likely to be male, Hispanic/Latino or Asian/Pacific Islander, students, highly educated, higher income, and COVID-vaccinated.

The manipulation created the intended experiential contrast. Understanding Chat participants reported substantially higher feelings of understanding (median = 13.0) compared to Deflective Chat (median = 2.0; Cliff's $d$ = −0.318, $p$ < .001), confirming successful manipulation. Within the structural equation model (SEM), FUM showed direct effects on article credibility (pro-vaccination: β = 0.41, 95 % CI [0.30, 0.52], $p$ < .001; anti-vaccination: β = 0.23, 95 % CI [0.12, 0.34], $p$ < .001) and indirect effects on vaccination intentions mediated through credibility (pro-vaccination: β = −0.07, 95 % CI [–0.13, −0.01], $p$ = .017; anti-vaccination: β = 0.09, 95 % CI [0.03, 0.15], $p$ = .006). While these paths from FUM represent correlative associations rather than causal experimental effects, they demonstrate the relationship between feeling understood and openness to opposing information within our theoretical model.

Turning to the experimental causal effects, the SEM revealed the hypothesised pathways from Understanding Chat condition operated as predicted but with crucial group differences (Figs. 3 and 4). In both groups, Understanding Chat influenced article credibility through FUM (indirect effect), though these effects were small and total effects on credibility were not significant in either group.

### 3.3. Study 2

Study 2 ($N$ = 752) extended the research to climate change scepticism and tested four conversational strategies. Climate sceptics were randomly assigned to: Neutral Chat ($n$ = 207; generic questions on random topics), Corrective Chat ($n$ = 165; respectfully countering climate misconceptions with scientific facts), Understanding Chat ($n$ = 181; exploring and acknowledging climate views without correction), or Friendly Chat ($n$ = 199; warm conversation on positive unrelated topics).

Conditions created different experiences ($\eta^2$ = 0.147, $p$ < .001). Friendly Chat produced the highest feelings of understanding (median = 16.0), Understanding Chat moderate (median = 13.0), Neutral Chat low (median = 4.0), and Corrective Chat lowest (median = 0.0). All comparisons were significant ($p$s < 0.003) except Neutral-Corrective. While conditions did not differ on article credibility ($p$ = .432), they did on voting intentions ($\eta^2$ = 0.017, $p$ = .005), with Friendly and Understanding Chats showing higher intentions than Neutral Chat.

Within the structural equation model, FUM again showed strong correlative pathways to both article credibility (β = 0.30, 95 % CI [0.23, 0.38], $p$ < .001) and voting intention, both indirectly through credibility (β = 0.15, 95 % CI [0.11, 0.19], $p$ < .001) and directly (β = 0.24, 95 % CI [0.17, 0.32], $p$ < .001), yielding a substantial total effect (β = 0.40, 95 % CI [0.32, 0.47], $p$ < .001). Unlike Study 1, baseline climate scepticism predicted feelings of understanding (β = −0.16, 95 % CI [–0.23, −0.09], $p$ < .001), with stronger sceptics reporting lower FUM (Fig. 5).

Examining experimental causal effects revealed multiple pathways to openness (Fig. 5). Understanding Chat produced the most consistent pattern: positive though marginal indirect effects on voting intention through the complete pathway (β = 0.02, 95 % CI [0.008, 0.04], $p$ = .002), significant total indirect effects (β = 0.06, 95 % CI [0.007, 0.11], $p$ = .026), and detectable total effects (β = 0.09, 95 % CI [0.01, 0.16], $p$ = .027). For credibility, Understanding Chat showed positive indirect effects through FUM (β = 0.05, 95 % CI [0.02, 0.07], $p$ = .002), though no net effect on credibility (β = 0.04, 95 % CI [–0.034, 0.11], $p$ = .316).

Friendly Chat mirrored and exceeded Understanding Chat's effects on voting intention, showing stronger indirect effects (β = 0.06, 95 % CI [0.04, 0.08], $p$ < .001) and total effects (β = 0.16, 95 % CI [0.08, 0.23], $p$ < .001). However, it exhibited a suppression effect (Field et al., 2012; Howell, 2010) for credibility: while showing positive indirect effects through FUM (β = 0.12, 95 % CI [0.08, 0.16], $p$ < .001), it simultaneously had a direct negative effect on credibility. These opposing forces cancelled out (total effect: β = 0.05, 95 % CI [–0.02, 0.12], $p$ = .163), with the negative direct effect only visible when controlling for FUM.

Corrective Chat showed marginal negative indirect pathways to both credibility (β = −0.03, 95 % CI [–0.06, −0.006], $p$ = .015) and voting

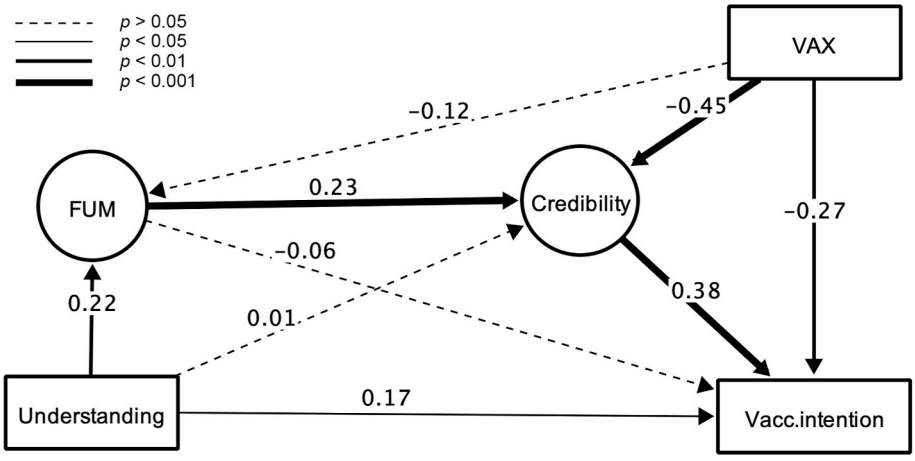


**Fig. 3.** Structural equation model with standardised coefficients for the anti-vaccination group (study 1).
*Note.* $N$ = 459. SEM1B; The Anti-Vaccination group read an article in favour of vaccination. Understanding = Understanding Chat condition; FUM = Feelings of (mis) understanding (higher = more understanding); Credibility = Perceived credibility of opposing article; Vacc.Intention = Intention to vaccinate fictitious child; VAX = Vaccination attitudes (higher = more sceptical). Model included demographic and procedural controls; $\chi^2$ = 904.03 ($df$ = 640, $p$ < .001), CFI = 0.995, TLI = 0.997, RMSEA = 0.042 (90 % CI = 0.036–0.049), SRMR = 0.063 (robust/scaled indices are reported); See SR Table S4.41 for full model details and fit indices.

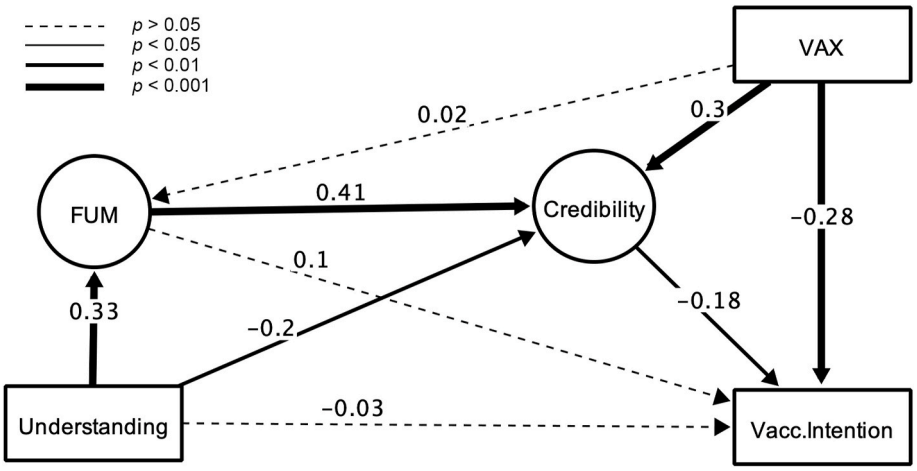


**Fig. 4.** Structural equation model with standardised coefficients for the pro-vaccination group (study 1).
*Note.* SEM1B; The Pro-Vaccination group read an article opposed to vaccination. Fig. 3 notes apply.

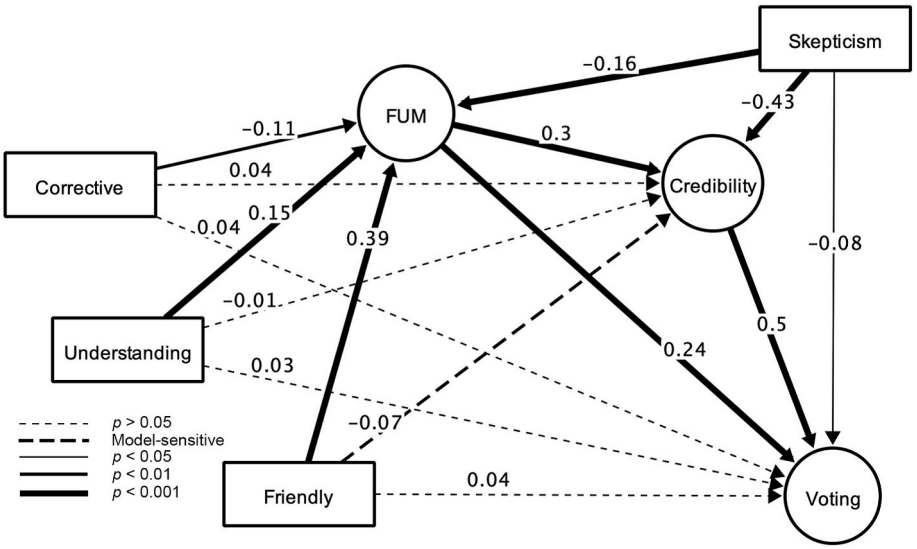


**Fig. 5.** Structural equation model with standardised coefficients (study 2).
*Note.* $N = 739$; SEM2B; Corrective = Corrective Chat; Understanding = Understanding Chat; Friendly = Friendly Chat; FUM = Feelings of (mis)understanding (higher = more understanding); Credibility = Perceived credibility of opposing article; Voting = Intention to vote for the political representative who authored the article; Scepticism = Climate change scepticism (higher = more sceptical). Model included demographic and procedural controls. $R^2$ values: FUM = 0.179, Credibility = 0.343, Voting intention = 0.442; $\chi^2 = 1552.36$ ($df = 476$, $p < .001$), CFI = 0.940, TLI = 0.934, RMSEA = 0.057 (90 % CI = 0.045–0.052), SRMR = 0.061 (robust/scaled indices are reported); See SR Table S5.43 for full model details.

intention (β = −0.02, 95 % CI [–0.03, −0.003], $p = .015$), with no significant total effects on either outcome.

The conversation length effect emerged earlier and more pronounced than Study 1. Conversations exceeding seven turns associated with decreased feelings of understanding (β = −0.11, 95 % CI [–0.18, −0.04], $p = .002$; Fig. 6).

The hierarchy of effectiveness for inducing feelings of understanding was: Friendly Chat (strongest but with credibility suppression), Understanding Chat (moderate with clean pathways), Neutral Chat (baseline), and Corrective Chat (counterproductive for immediate openness).

### 3.4. Study 3

Study 3 examined persistence by re-measuring outcomes 60 days after Study 2. Of 752 eligible CloudConnect participants, 378 (50.3 %) completed follow-up. Non-returning participants differed systematically—more likely employed, from incnome extremes, older, more climate-sceptical, and from Corrective Chat (all $ps < 0.05$). Despite differential retention, the final sample maintained balance across conditions: Neutral ($n = 91$), Corrective ($n = 99$), Understanding ($n = 89$), and Friendly ($n = 99$), with no baseline climate scepticism differences between conditions. Average follow-up was 60.0 days (Supplementary Results S6.2–6.3).

Re-analysis of Study 2 variables within the retained sample confirmed original patterns with increased effect sizes in this higher-quality CloudConnect subsample. Though Corrective Chat's negative effect on FUM became marginally non-significant ($p = .051$–$0.146$), likely due to reduced power.

The most consistent finding was the persistence of FUM's correlative pathways. Even 60 days later, baseline feelings of understanding predicted both reduced climate scepticism (β = −0.20, 95 % CI [–0.31, −0.08], $p = .001$) and increased voting intentions (β = 0.33, 95 % CI [0.22, 0.45], $p < .001$), controlling for demographics and baseline scepticism. While FUM's effects were partially mediated by credibility,

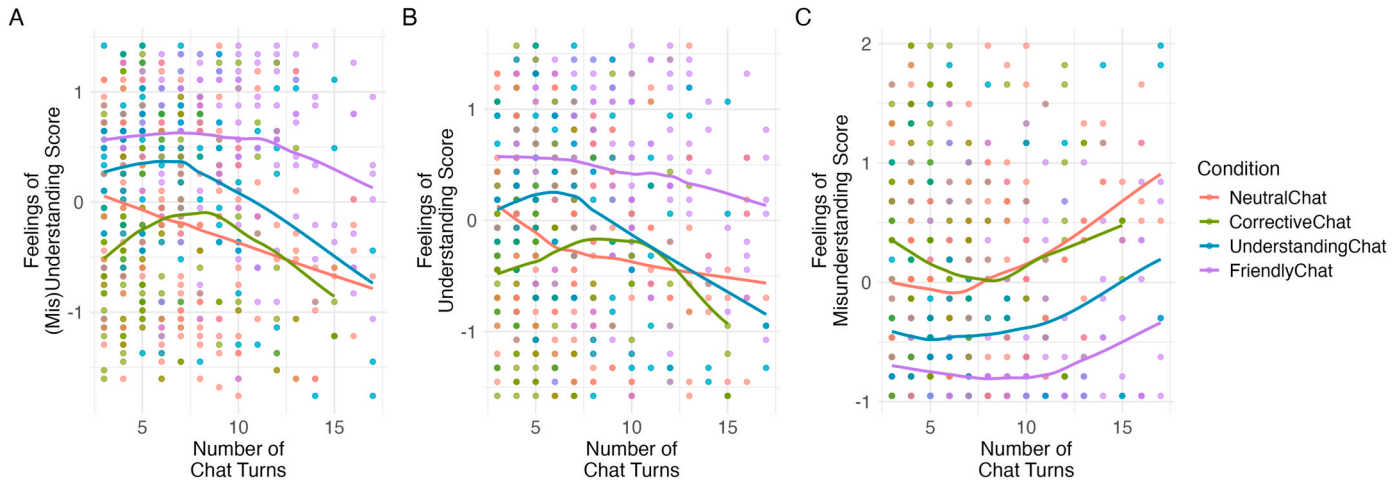


**Fig. 6.** Relationships among chat conditions, number of chat turns, and feelings of (Mis)Understanding (study 2).
*Note.* Individual participant observations colored by condition with LOESS-smoothed trend lines. Data filtered to 5th–95th percentile for clarity. Feelings of (mis) understanding represents combined score from both scale dimensions.

considerably higher total effects suggest durable influence beyond the measured pathway. The establishment of temporal precedence, with FUM measured 60 days prior to outcomes, strengthens evidence for a potential causal relationship (Hayes & Little, 2022).

Experimental conditions showed evolving patterns over time (Fig. 7). For voting intentions (new scenario without reference to Study 2), Understanding Chat maintained positive indirect effects (β = 0.07, 95 % CI [0.02, 0.12], $p$ = .012), including the complete, though marginal, two-link mediation through FUM and credibility (β = 0.01, 95 % CI [0.001, 0.03], $p$ = .039). While Understanding Chat's total effect was positive but unreliable (β = 0.08, 95 % CI [–0.04, 0.20], $p$ = .208), likely due to insufficient power.

Friendly Chat still showed stronger indirect effects on voting intentions (β = 0.12, 95 % CI [0.05, 0.18], $p$ = .001) including through the complete pathway (β = 0.04, 95 % CI [0.02, 0.07], $p$ = .001). However, these positive indirect effects were completely offset by a paradoxical direct negative effect (See Fig. 9 for a comparison of short- and long-term effects), resulting in no net change (β = −0.01, 95 % CI [–0.14, 0.11], $p$ = .812).

Surprisingly, Corrective Chat, which showed no immediate benefits in Study 2, emerged with a direct positive effect on voting intentions exceeding any other condition (β = 0.13, 95 % CI [0.02, 0.24], $p$ = .025), though its total effect remained marginally non-significant (β = 0.12, 95 % CI [–0.003, 0.25], $p$ = .055). Kruskal-Wallis tests confirmed condition differences ($\chi^2$ = 10.91, d. f. = 3, $p$ = .012, $\eta^2$ = 0.029), with pairwise comparisons showing reliable differences ($Z$ = −2.68, $p$ = .044) only between Corrective and Friendly conditions (medians: Neutral = 16, Corrective = 19, Understanding = 19, Friendly = 16).

For climate scepticism reduction, patterns aligned with voting intentions though most effects fell short of conventional thresholds. Understanding Chat and Friendly Chat showed detectable indirect pathways through FUM and credibility, with Friendly producing greater reductions through this route. However, total effects diverged: Friendly Chat's was negligible (β = −0.02, 95 % CI [–0.10, 0.06], $p$ = .593), while Understanding Chat's was larger though marginally non-significant (β = −0.07, 95 % CI [–0.15, 0.01], $p$ = .087). Corrective Chat showed a direct negative effect on scepticism (β = −0.09, 95 % CI [–0.19, 0.000], $p$ = .053), reliable without control variables.

The longitudinal findings revealed that immediate feelings of understanding created lasting effects on both attitudes and intentions, while conversational strategies showed complex temporal dynamics—with corrective approaches potentially requiring time for defensive reactions to subside before benefits emerge.

### 3.5. *Commentary on effect size*

The hypothesised mediation pathway—from experimental conditions through FUM and credibility to behavioural intentions—showed small but consistent effects (β = 0.02–0.06, except for Pro-Vaccination participants, Fig. 8). Full mediation with two links is challenging to detect as each link must function and compound multiplicatively (Fritz & MacKinnon, 2007; Hayes & Little, 2022; Rucker et al., 2011). While

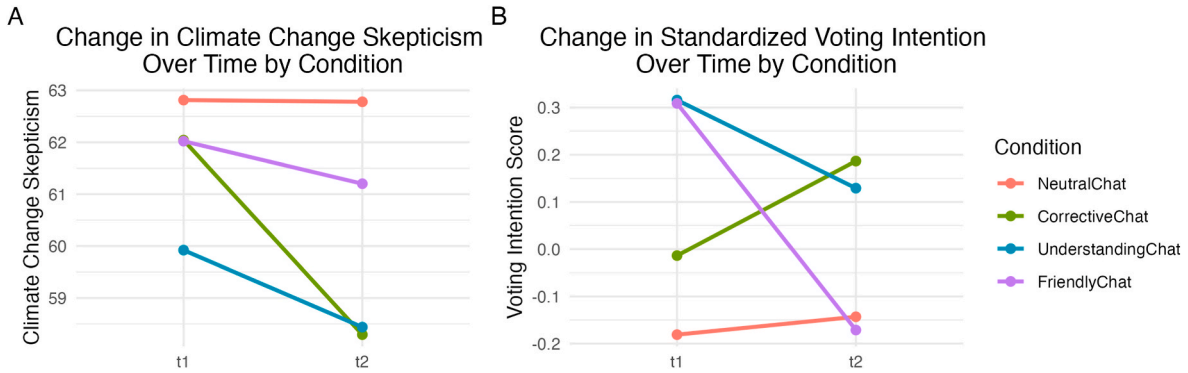


**Fig. 7.** Changes in climate scepticism and voting intention from study 2 to study 3.
*Note.* (A) Change in mean climate change scepticism from baseline (Study 2) to 60-day follow-up across four conditions. (B) Change in mean voting intention; scores standardised due to different scenario framing between timepoints (Study 2: chat partner as candidate; Study 3: moderate politician with climate priorities).

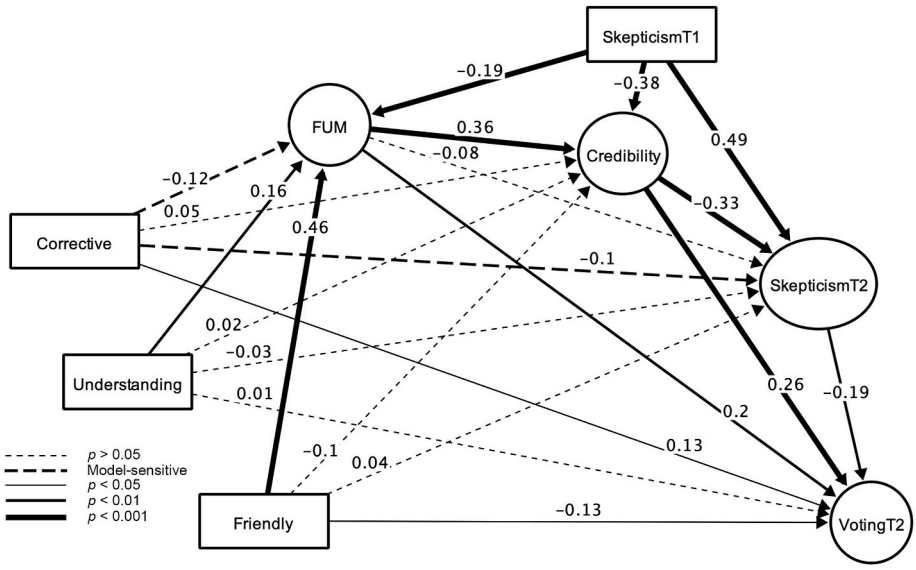


**Fig. 8.** Structural equation model with standardised coefficients (study 3).
*Note.* $N = 374$; SEM3B; Corrective = Corrective Chat; Understanding = Understanding Chat; Friendly = Friendly Chat; FUM = Feelings of (mis)understanding (higher = more understanding); Credibility = Perceived credibility of opposing article; VotingT2 = Voting intention at follow-up; ScepticismT1/T2 = Climate change scepticism at baseline/60-day follow-up. - = 374; $R^2$ values: FUM = 0.278, Credibility = 0.373, Scepticism T2 = 0.533, Voting T2 = 0.253. $\chi^2 = 1740.49$ ($df = 845$, $p < .001$), CFI = 0.905, TLI = 0.898, RMSEA = 0.054 (90 % CI = 0.051–0.058), SRMR = 0.057 (robust/scaled indices are reported); See SR Table S6.26 for full details.

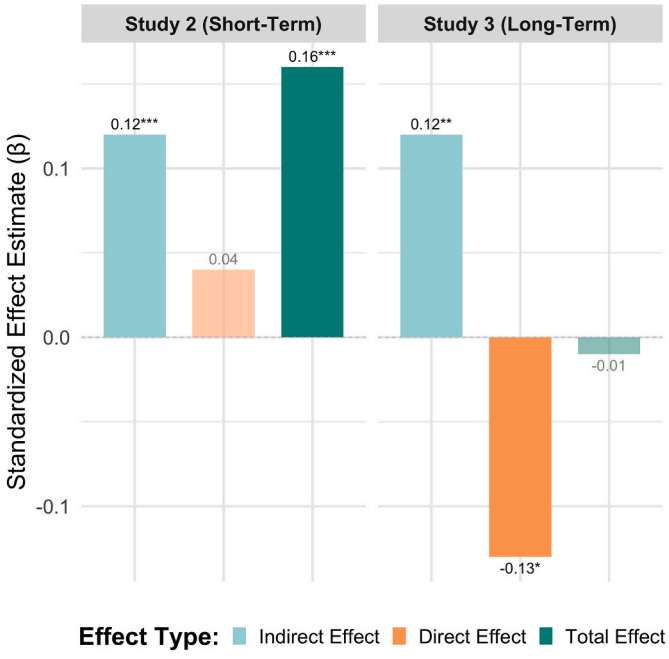


**Fig. 9.** Competing mechanisms of friendly chat on voting intention.
*Note.* The chart displays standardised effect estimates (β) from the structural equation models, comparing the effects of the Friendly Chat condition (relative to the Neutral Chat control) on voting intention in Study 2 (Short-Term) and Study 3 (Long-Term). "Indirect Effect" represents the sum of all mediated pathways (e.g., via FUM); "Direct Effect" is the unmediated path from the condition to the outcome; "Total Effect" is the sum of all direct and indirect paths. The figure highlights how the positive indirect effect is supported by a non-significant positive direct effect in the short term (Study 2) but is cancelled by an emergent, significant negative direct effect in the long term (Study 3). Bars with reduced opacity represent non-significant paths. *$p < .05$, **$p < .01$, **$p < .001$.

the experimental mediation effects are modest, their theoretical and practical importance is substantial for three reasons. First, they were produced by a minimal, 6-min intervention targeting deeply-held beliefs. Second, they demonstrated persistence over a 60-day follow-up, a high bar for brief interventions. Third, and most critically, these small effects were consistently observed across multiple studies, a pattern that meets the criteria for theoretical relevance and suggests a robust, replicable phenomenon (Götz et al., 2022; Primbs et al., 2023). Fig. 10 presents a summary of effects across all studies.

## 4. Discussion

This research employed novel methodology, combining experimentally controlled LLM-mediated conversations with a longitudinal 60-day follow-up to track the persistence and evolution of persuasive effects. It demonstrated that brief AI conversations can enhance openness to opposing information through induced feelings of understanding, with effects persisting 60 days post-interaction but varying markedly by conversational strategy. While Understanding Chat fostered openness through expected pathways, Friendly Chat revealed competing mechanisms and Corrective Chat showed delayed benefits.

### 4.1. *Feelings of understanding*

In line with the hypothesis strong correlational pathways emerged from feelings of understanding (FUM) to perceived credibility of opposing views and counter-attitudinal behavioural intentions across all studies. However, the FUM scale's sensitivity to non-topical positive conversation—with Friendly Chat producing twice the effect of Understanding Chat—suggests it captures general affective experience rather than a more complex affective-cognitive construct, as others have suggested (Grice, 1997; Schrodt, 2003; Schrodt & Finn, 2011) and why this scale was chosen. Without baseline mood assessment, correlational effects cannot be attributed exclusively to the manipulated experience of *feeling understood*.

However, the durability of these pathways at 60-day follow-up strengthens evidence for potential causality (Hayes & Little, 2022) and supports the presented ELM (Petty & Cacioppo, 1986) mechanisms whereby positive affective states from feeling understood (Morelli et al.,

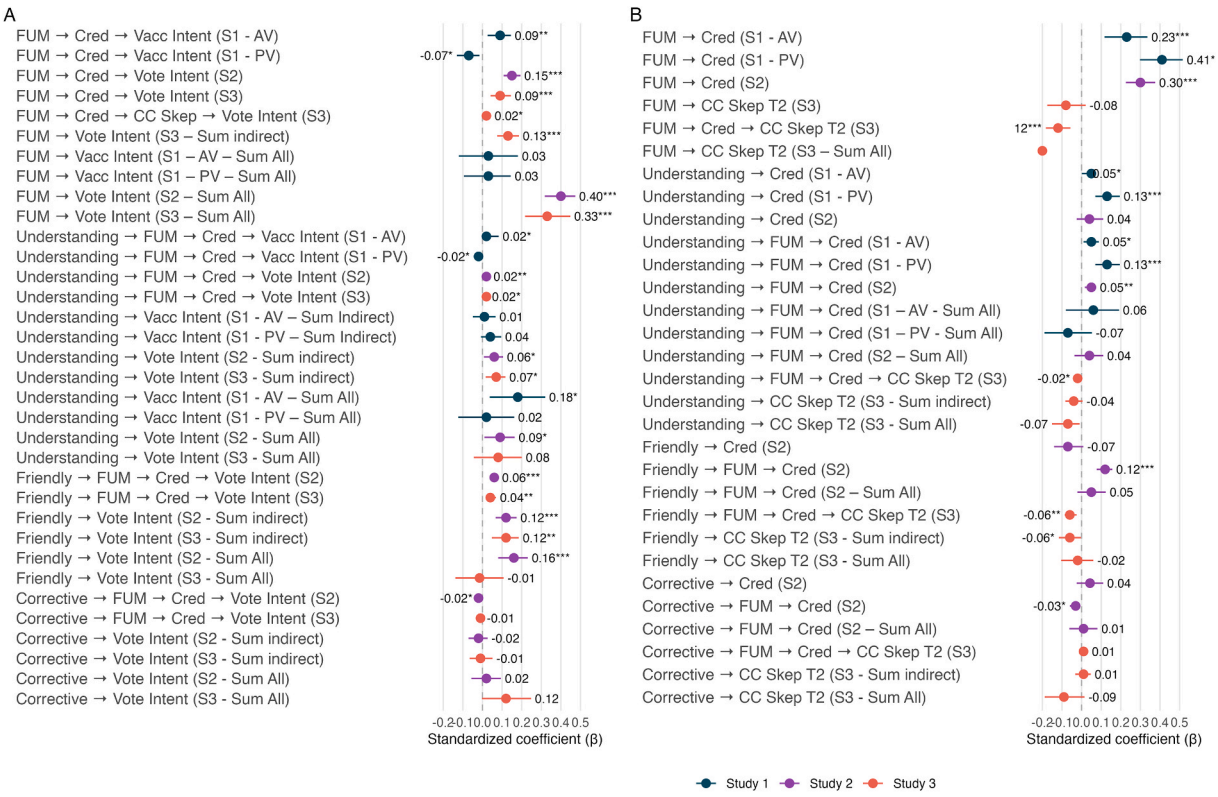


**Fig. 10.** Indirect Pathways from Experimental Conditions to Outcomes Across three Studies.
*Note.* Forest plots showing standardised indirect and total effects from structural equation models. (A) Effects on behavioural intentions (vaccination intention in Study 1, voting intention in Studies 2–3). (B) Effects on credibility and climate change scepticism at follow-up (Study 3). Points represent standardised coefficients (β) with 95 % confidence intervals relative to control conditions (Deflective Chat in Study 1, Neutral Chat in Studies 2–3). AV = anti-vaccination group; PV = pro-vaccination group; FUM = Feelings of (mis)understanding; CC Skep T2 = Climate change scepticism at 60-day follow-up; Sum Indirect = sum of indirect effects; Sum All = total effect. *$p < .05$, **$p < .01$, **$p < .001$.

2014) could enhance openness through multiple routes: increasing processing motivation and ability, influencing elaboration bias favourably, or serving as peripheral cues signalling source trustworthiness.

### 4.2. *Competing mechanisms and the complexity of AI-mediated understanding*

The hypothesis received further support through causal relationships for counter-attitudinal behavioural intentions. The Understanding Chat predicted vaccination intentions in Anti-Vaccination participants (Pilot, Study 1) and voting intentions (Study 2), with weaker evidence 60 days later in Study 3. The Friendly Chat showed stronger voting intention effects than Understanding Chat in Study 2 but no effect in Study 3.

While behavioural intention effects were consistently mediated by FUM, full mediation through both FUM and credibility yielded small effect sizes despite being clearly non-random, suggesting additional pathways which require further research. Critically, examining credibility revealed counteracting processes: mediated effects through FUM were positive, but total effects were smaller or absent. The more nuanced Understanding Chat showed mixed evidence for this, which was mostly only indicated and clearly observed in the majority opinion pro-vaccination group in Study 1. However, the Friendly Chat demonstrated clear suppression (Field et al., 2012; Howell, 2010) effects—positive mediated pathways increased credibility and intentions while direct negative effects cancelled these gains. This "driving with brakes on" pattern emerged weakly in Study 2 but crystallised in Study 3, ultimately returning outcomes to Neutral Chat levels.

This may represent the most interesting finding: humans could potentially operate with two systems when encountering AI-mediated understanding. One system may respond automatically to displays of understanding regardless of source authenticity, while another possibly simultaneously detects and resists artificiality—An *authenticity paradox* which demands further investigation.

**Short-vs. Long-term Effects:** Crucially, in the short term, the competing mechanisms were only observed for the perceived credibility of opposing views; the voting intention remained untouched, with negative effects manifesting only in the long term. However, the temporary openness to voting for a counter-attitudinal politician might also be explained by the political representative being framed as the chat partner exclusively in Study 2 and Broaden-and-Build Theory, which posits that positive emotions can temporarily increase openness by broadening thought-action repertoires (Fredrickson, 2001; Fredrickson & Branigan, 2005).

**AI Suspicion:** These competing mechanisms align with evidence of human reactions to AI: Well-intentioned AI interactions can elicit suspicion when viewed as artificial or manipulative (Yin et al., 2024), users increasingly prefer less overtly human-like LLMs (Cheng et al., 2025), and AI perceived as lacking genuine empathy (Wygnańska, 2023) or reciprocity (Brandtzaeg et al., 2022) creates unease. These findings could explain our experimental conditions' dual nature: triggering evolutionarily ingrained positive reactions while simultaneously activating negative responses through perceived inauthenticity.

In addition to these competing forces, Friendly Chat's effects may have been more complex than simple cancellation. Temporary effects suggest peripheral route processing (Petty et al., 1993, 1995), and others have noted good mood reduces systematic processing (Schwarz & Bless, 1991; Worth & Mackie, 1987). However, durable correlational effects and positive indirect effects at follow-up higher than from Understanding Chat indicate that high FUM scores may have enabled lasting changes beyond what positive affect alone would predict. Positive affect

or an activated reward system may have freed cognitive resources or enhanced source likability sufficiently for thoughtful engagement with opposing information.

Conversely, Friendly Chat may have produced durable negative effects resembling the Pilot Study, where concealing AI identity undermined efficacy. Through an ELM lens (Petty & Cacioppo, 1986), off-topic positive interaction before a persuasive task could seem mismatched, creating suspicion. While initial positive affect might have enabled elaboration, perceived manipulation could have damaged source trust, prompted negatively biased processing, or acted as a negative peripheral cue.

**Corrective Chat Surprise:** The Corrective Chat warrants special attention. Despite respectful delivery, factual corrections produced lower feelings of understanding and no immediate effects. However, Study 3 revealed this as the only condition with positive *direct* effects on long-term voting intentions and trends toward reduced climate scepticism. These effects occurred directly, without mediation through FUM or credibility, suggesting respectfully delivered facts may be processed and integrated over time despite initial resistance. This supports previous findings of durable chatbot correction effects (Costello et al., 2024). While resembling the sleeper effect, absent immediate effects make it difficult to establish whether necessary conditions were met (Kumkale & Albarracín, 2004), though delayed positive intentions suggest deeper cognitive processing occurred.

### 4.3. The influence of pre-existing stances and participant demographics

Pro- and anti-vaccination groups showed differential effects in Studies 1 and 2. While the experimental pathway (Understanding Chat→FUM→credibility→intention) appeared in both groups, effect sizes differed markedly. Pro-vaccination participants showed 50 % higher effects from condition to FUM and nearly double the effect from FUM to credibility, yet only half the magnitude from credibility to intention. Combined with negative direct effects on credibility, this yielded no detectable total effects on vaccination intention. These dynamics suggest persuasion processes vary by attitudinal position, consistent with research showing minorities benefit more from feeling heard than majorities (Bruneau & Saxe, 2012).

Consistent demographic patterns emerged for women and Black/African American participants who rated opposing articles as more credible across studies (except pro-vaccination participants in Study 1). Less consistently, age was positively associated with feelings of understanding but inversely with credibility of opposing information, while higher education predicted lower feelings of understanding. Current literature offers little explanation—the Receptiveness to Opposing Views scale found no gender differences (Minson et al., 2020), and cultural differences in feeling understood concern different aspects of self-perception (Oishi et al., 2010). These patterns require targeted future investigation.

### 4.4. Eliciting "feeling understood" via AI

Studies 1 and 2 confirmed that AI chatbots effectively induced feelings of being understood, aligning with the Computers Are Social Actors paradigm (Reeves & Nass, 1996). This extends evidence that individuals form relational responses to AI, from companionate bonds to perceiving human-like minds (Brandtzaeg et al., 2022; De Freitas et al., 2024; Lee & Hahn, 2024; Wygnańska, 2023), supported by LLMs' demonstrated sophisticated emotional understanding (Schlegel et al., 2025). Yet humans remain discerning, as previous findings outlined above and the competing mechanisms revealed. FUM scores peaked around ten conversation turns before declining, possibly suggesting extended interactions exposed limitations.

### 4.5. Practical implications

The findings have implications for countering misinformation and societal division central to cognitive warfare. On a professional level, practitioners in conflict-prone occupational fields—teachers, doctors, police, governments, politicians—could prioritize authentic understanding and acknowledgment without fearing legitimization of extremes, as neither this nor past research appears to support such a mechanism.

**Who is Right?** The Understanding Chat's success may reflect the epistemic value of abandoning truth dominance—a core scientific principle (Popper, 1984) that society could fearlessly embrace, since not only this research shows genuine understanding yields precious benefits: increased openness and intellectual humility (Itzchakov et al., 2024; Minson & Chen, 2022), reduced prejudice and political separatism (Livingstone, Fernández Rodríguez, & Rothers, 2020), greater cross-difference engagement (Livingstone, Fernández Rodríguez, & Rothers, 2020; Yeomans et al., 2020), and strengthened ideological-crossing bonds (Reschke et al., 2020). Thus, educational curricula should expand beyond debunking to include receptive communication training and foster acceptance of contrary opinions.

**The Bitter Pill:** artificial chatbot conversations worked despite their disingenuous nature. This demonstrates abuse potential for AI and induction of positive affect in general. These responses may be evolutionarily ingrained—part of us appears to respond to artificial displays of understanding, unable to distinguish genuine from performed empathy. Beyond AI providers, malicious actors—influencers, politicians, marketers—may exploit induced feelings of understanding. Demagogues have long used "I understand you" tactics to capture those feeling unheard (Engesser et al., 2017; Kaltwasser et al., 2017; Şimşek, 2024); AI now enables unprecedented scaling. However, the observed negative effects in our studies indicate latent resilience against such manipulation. This points to the potential value of education in recognizing weaponized performance and enhance our ability to distinguish authentic from manipulative understanding.

Vulnerability to such manipulation may stem from perceived or real unmet needs for understanding, with stronger effects documented in minority groups (Bruneau & Saxe, 2012). Concerns about health, economic, or cultural changes should be met with acknowledgment rather than dismissal, as otherwise individuals may become receptive to any source offering validation—including populist movements. Current societal divisions may partially reflect fundamental needs for understanding and belonging (Baumeister & Leary, 1995; Deci & Ryan, 1987; Swann Jr, 1990), and fostering understanding of opposing opinions pulls the rug out from under division-based societal attacks. What is more dangerous—Differences in our truths or the divisions we let them create?

### 4.6. Limitations

Several limitations constrain interpretation. First, "feeling understood" lacks uniform definition or measurement (Lun et al., 2008; Morelli et al., 2014; Reis et al., 2017). The FUM scale's sensitivity to non-topical positive conversation (i.e., the Friendly Chat condition) suggests it may measure general affective states, rather than a nuanced, cognitive appraisal of genuine, topic-specific understanding, and does not distinguish it from mere positive affect or reward system activation (Oishi et al., 2010; Reis & Gable, 2015).

Second, while the ELM framework provided theoretical grounding, it cannot be definitely identified which specific pathways activated in Understanding Chat or Friendly Chat—enhanced motivation, increased ability, altered processing, or peripheral cues. Though persistence and credibility mediation suggest central route processing (Petty et al., 1995), precise mechanisms remain speculative.

Third, operationalising openness as credibility ratings and counter-attitudinal intentions captures only one dimension, potentially missing

shifts in cognitive flexibility or epistemic attitudes that broader conceptualizations include (Itzchakov et al., 2024; Minson et al., 2020; Reis et al., 2017).

Fourth, samples were non-representative with high exclusion rates, particularly affecting men and lower-educated participants. Differing pro/anti-vaccination dynamics suggest persuasion varies by attitudinal position, limiting generalizability.

Fifth, artificial experimental contexts—single interactions followed by prescribed articles—produced small effect sizes. AI understanding may differ fundamentally from human understanding, limiting theoretical conclusions about interpersonal persuasion.

Sixth, some structural equation models required adaptations for convergence, Study 3 was underpowered, resulting in a model with poor overall fit and insufficient sample size for the model's complexity. Therefore, all findings regarding the persistence of specific pathways at 60 days must be interpreted with significant caution and require replication.

Seventh, while prompts and fine-tuning data are provided, the inherent stochastic nature of LLMs means that conversational content inevitably varied between participants. This introduces variance that limits the feasibility of a strict replication.

Finally, Study 2 and 3 occurred shortly after Trump's re-election potentially affecting minority/majority dynamics, testing only two topics with single exposures limits domain insights, and findings need replication.

### 4.7. Future research

Future research should distinguish whether effects stem from reward system activation, general positive affect, or genuine feelings of understanding through targeted experimental manipulation. It has previously been found that different kinds of positive emotions affect processing differently (Griskevicius et al., 2010); feeling understood might present its own mechanisms. Identifying which ELM pathways activate—motivation, ability, processing valence, or peripheral cues—requires systematic investigation. The boundaries between machine-made, genuine, and strategic understanding warrant examination, as does developing standardised measures for both feelings of understanding and openness constructs.

Unexpected findings merit exploration. Friendly Chat's negative direct effects may represent a novel inauthenticity-driven backfire. Critical questions include when authenticity detection overrides automatic understanding responses, why short term effects differed between credibility and behavioural intentions, and whether mainstream-aligned individuals benefit less from validation they already receive.

Methodological improvements include representative samples by removing chat/reading barriers through voice-based interactions and video presentations. Testing diverse topics beyond vaccination/climate, examining minority/majority configurations, and longitudinal designs with baseline mood controls would enhance generalizability and causal inference. A hybrid approach combining Understanding Chat with carefully integrated corrections should be explored.

### 4.8. Conclusion

Most remarkably, any persistence after 60 days from a single 6-min interaction is noteworthy. Findings suggest that feeling understood increases openness to opposing information and fosters counterattitudinal behaviour both in the short and long term. Divergent temporal dynamics emphasize the critical importance of longitudinal assessment—immediately effective interventions may disappear while seemingly ineffective approaches plant seeds for future change.

This work reveals both promise and peril: Feeling understood can bridge divides but is vulnerable to manipulation. Fortunately, a competing detection system may offer potential protection against inauthentic understanding. It began confronting cognitive warfare—deliberate manipulation fracturing societies. It concludes with an unexpectedly simple insight: in an era of sophisticated disinformation and algorithmic polarization, the path forward might be as basic as making people feel heard. Perhaps countering these attacks requires simply genuine acceptance and friendliness.

## Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used Google Gemini and Claude (Anthropic) in order to improve prose, sentence structure, and clarity of expression. After using these tools, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the published article.

## Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. Recruitment costs were covered by the University of Bern.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chb.2025.108870.

## Data availability

All anonymous data and code are available at https://osf.io/tn4cq/
Preregistration Pilot: https://aspredicted.org/jq4c-mhvh.pdf
Preregistration Study 1: https://aspredicted.org/7m9n-zm8x.pdf
Preregistration Study 2: https://aspredicted.org/kdr9-tx67.pdf

## References

Agresti, A. (2010). *Analysis of ordinal categorical data* (2nd ed.). Wiley.

Andrade, C. (2018). Internal, external, and ecological validity in research design, conduct, and evaluation. *Indian Journal of Psychological Medicine, 40*(5), 498–499. https://doi.org/10.4103/IJPSYM.IJPSYM_334_18

Ballew, M. T., Leiserowitz, A., Roser-Renouf, C., Rosenthal, S. A., Kotcher, J. E., Marlon, J. R., Lyon, E., Goldberg, M. H., & Maibach, E. W. (2019). Climate change in the American mind: Data, tools, and trends. *Environment: Science and Policy for Sustainable Development, 61*(3), 4–18. https://doi.org/10.1080/00139157.2019.1589300

Basol, M., Roozenbeek, J., & Van Der Linden, S. (2020). Good news about bad news: Gamified inoculation boosts confidence and cognitive immunity against fake news. *Journal of Cognition, 3*(1), 2. https://doi.org/10.5334/joc.91

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin, 117*(3), 497–529. https://doi.org/10.1037/0033-2909.117.3.497

Begum, T., Efstathiou, N., Bailey, C., & Guo, P. (2024). Cultural and social attitudes towards COVID-19 vaccination and factors associated with vaccine acceptance in adults across the globe: A systematic review. *Vaccine.* , Article 125993. https://doi.org/10.1016/j.vaccine.2024.05.041

Bentler, P. M., & Chou, C.-P. (1987). Practical issues in structural modeling. *Sociological Methods & Research, 16*(1), 78–117. https://doi.org/10.1177/0049124187016001004

Bertlich, T., Bräscher, A., Germer, S., Witthöft, M., & Imhoff, R. (2025). Owners of a conspiratorial heart? Investigating the longitudinal relationship between loneliness and conspiracy beliefs. *British Journal of Social Psychology, 64*(2), Article e12865. https://doi.org/10.1111/bjso.12865

Bertrams, A., & Krispenz, A. (2025). Antisemitism as a dark-ego vehicle. *Current Psychology, 44*(1), 676–692. https://doi.org/10.1007/s12144-024-07120-z

Betsch, C., & Sachse, K. (2013). Debunking vaccination myths: Strong risk negations can increase perceived vaccination risks. *Health Psychology: Official Journal of the Division of Health Psychology, 32*(2), 146–155. https://doi.org/10.1037/a0027387. American Psychological Association.

Biddlestone, M., Green, R., Douglas, K., Azevedo, F., Sutton, R. M., & Cichocka, A. (2022). Reasons to believe: A systematic review and meta-analytic synthesis of the motives associated with conspiracy beliefs. https://doi.org/10.31234/osf.io/rxjqc.
Bloch-Atefi, A. (2025). Balancing ethics and opportunities: The role of AI in psychotherapy and counselling. *Psychotherapy and Counselling Journal of Australia*. https://doi.org/10.59158/001c.129884
Brandtzaeg, P. B., Skjuve, M., & Følstad, A. (2022). My AI friend: How users of a social chatbot understand their Human–AI friendship. *Human Communication Research, 48*(3), 404–429. https://doi.org/10.1093/hcr/hqac008
Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. The Guilford Press.
Bruneau, E. G., & Saxe, R. (2012). The power of being heard: The benefits of 'perspective-giving' in the context of intergroup conflict. *Journal of Experimental Social Psychology, 48*(4), 855–866. https://doi.org/10.1016/j.jesp.2012.02.017
Burgoon, J. K. (1993). Interpersonal expectations, expectancy violations, and emotional communication. *Journal of Language and Social Psychology, 12*(1–2), 30–48. https://doi.org/10.1177/0261927X93121003
Cahn, D. D., & Shulman, G. M. (1984). The perceived understanding instrument. *Communication Research Reports, 1*(1), 122–125. https://doi.org/10.1080/17464099.1984.12289906
Chan, H.-W., Wang, X., Tam, K.-P., Hong, Y., & Huang, B. (2024). Hotter weather, less of a hoax? Testing the longitudinal association between experience of temperature anomalies and belief in climate change conspiracy theories. *Journal of Environmental Psychology, 98*, Article 102409. https://doi.org/10.1016/j.jenvp.2024.102409
Chandler, J., Rosenzweig, C., Moss, A. J., Robinson, J., & Litman, L. (2019). Online panels in social science research: Expanding sampling methods beyond mechanical turk. *Behavior Research Methods, 51*(5), 2022–2038. https://doi.org/10.3758/s13428-019-01273-7
Cheng, M., Yu, S., & Jurafsky, D. (2025). *HumT DumT: Measuring and controlling human-like language in LLMs* (version 1). *arXiv*. https://doi.org/10.48550/ARXIV.2502.13259
Claverie, B. (2024). "Cognitive Warfare" – Une guerre invisible qui s'attaque à notre pensée. In J.-F. Trinquecoste (Ed.), *Faut-il s'inquiéter ?* (pp. 89–115). Éditions de l'IAPTSEM. https://hal.science/hal-04586061.
Claverie, B., & Du Cluzel, F. (2022). "Cognitive Warfare": The Advent of the Concept of "Cognitics" in the field of warfare. In B. Claverie, B. Prébot, N. Buchler, & F. du Cluzel (Eds.), *Cognitive warfare: The future of cognitive dominance*. NATO Collaboration Support Office, 2, 1–7 https://hal.science/hal-03635889.
Cook, J., Ecker, U. K. H., Trecek-King, M., Schade, G., Jeffers-Tracy, K., Fessmann, J., Kim, S. C., Kinkead, D., Orr, M., Vraga, E., Roberts, K., & McDowell, J. (2023). The cranky uncle game—Combining humor and gamification to build student resilience against climate misinformation. *Environmental Education Research, 29*(4), 607–623. https://doi.org/10.1080/13504622.2022.2085671
Cosgrove, T. J., & Murphy, C. P. (2023). Narcissistic susceptibility to conspiracy beliefs exaggerated by education, reduced by cognitive reflection. *Frontiers in Psychology, 14*, Article 1164725. https://doi.org/10.3389/fpsyg.2023.1164725
Costello, T. H., Pennycook, G., & Rand, D. G. (2024). Durably reducing conspiracy beliefs through dialogues with AI. *Science, 385*(6714). https://doi.org/10.1126/science.adq1814. eadq1814.
De Freitas, J., Uğuralp, A. K., Uğuralp, Z., & Puntoni, S. (2024). AI companions reduce loneliness. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4893097
De Graaf, J. A., Stok, F. M., De Wit, J. B. F., & Bal, M. (2023). The climate change skepticism questionnaire: Validation of a measure to assess doubts regarding climate change. *Journal of Environmental Psychology, 89*. https://doi.org/10.1016/j.jenvp.2023.102068, 102068.
Deci, E. L., & Ryan, R. M. (1987). The support of autonomy and the control of behavior. *Journal of Personality and Social Psychology, 53*(6), 1024–1037. https://doi.org/10.1037/0022-3514.53.6.1024
Deppe, C., & Schaal, G. S. (2024). Cognitive warfare: A conceptual analysis of the NATO ACT cognitive warfare exploratory concept. *Frontiers in Big Data, 7*, Article 1452129. https://doi.org/10.3389/fdata.2024.1452129
DeVellis, R. F. (2017). *Scale development: Theory and applications* (4th ed.). SAGE.
Döring, N., & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-41089-5
Douglas, B. D., Ewell, P. J., & Brauer, M. (2023). Data quality in online human-subjects research: Comparisons between MTurk, Prolific, CloudResearch, Qualtrics, and SONA. *PLoS One, 18*(3), Article e0279720. https://doi.org/10.1371/journal.pone.0279720
Douglas, K. M., Sutton, R. M., Biddlestone, M., Green, R., & Toribio-Flórez, D. (2024). Engaging with conspiracy believers. *Review of Philosophy and Psychology*. https://doi.org/10.1007/s13164-024-00741-0
Douglas, K. M., Sutton, R. M., Callan, M. J., Dawtry, R. J., & Harvey, A. J. (2016). Someone is pulling the strings: Hypersensitive agency detection and belief in conspiracy theories. *Thinking & Reasoning, 22*(1), 57–77. https://doi.org/10.1080/13546783.2015.1051586
Douglas, K. M., Sutton, R. M., & Cichocka, A. (2017). The psychology of conspiracy theories. *Current Directions in Psychological Science, 26*(6), 538–542. https://doi.org/10.1177/0963721417718261
Douglas, K. M., Sutton, R. M., & Cichocka, A. (2019). Belief in conspiracy theories. In J. P. Forgas, & R. F. Baumeister (Eds.), *The social psychology of gullibility* (1st ed., pp. 61–76). Routledge. https://doi.org/10.4324/9780429203787-4.
Douglas, K. M., Sutton, R. M., Van Lissa, C. J., Stroebe, W., Kreienkamp, J., Agostini, M., Bélanger, J. J., Gützkow, B., Abakoumkin, G., Khaiyom, J. H. A., Ahmedi, V., Akkas, H., Almenara, C. A., Atta, M., Bagci, S. C., Basel, S., Berisha Kida, E., Bernardo, A. B. I., Buttrick, N. R., … Leander, N. P. (2023). Identifying important individual- and country-level predictors of conspiracy theorizing: A machine learning analysis. *European Journal of Social Psychology, 53*(6), 1191–1203. https://doi.org/10.1002/ejsp.2968
Dyrendal, A., Kennair, L. E. O., & Bendixen, M. (2021). Predictors of belief in conspiracy theory: The role of individual differences in schizotypal traits, paranormal beliefs, social dominance orientation, right wing authoritarianism and conspiracy mentality. *Personality and Individual Differences, 173*, Article 110645. https://doi.org/10.1016/j.paid.2021.110645
Ecker, U. K. H., Lewandowsky, S., Cook, J., Schmid, P., Fazio, L. K., Brashier, N., Kendeou, P., Vraga, E. K., & Amazeen, M. A. (2022). The psychological drivers of misinformation belief and its resistance to correction. *Nature Reviews Psychology, 1*(1), 13–29. https://doi.org/10.1038/s44159-021-00006-y
Engesser, S., Ernst, N., Esser, F., & Büchel, F. (2017). Populism and social media: How politicians spread a fragmented ideology. *Information, Communication & Society, 20*(8), 1109–1126. https://doi.org/10.1080/1369118X.2016.1207697
Feinberg, M., & Willer, R. (2013). The moral roots of environmental attitudes. *Psychological Science, 24*(1), 56–62. https://doi.org/10.1177/0956797612449177
Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. Sage.
Flanagin, A. J., & Metzger, M. J. (2000). Perceptions of internet information credibility. *Journalism & Mass Communication Quarterly, 77*(3), 515–540. https://doi.org/10.1177/107769900007700304
Fox, J., & Gambino, A. (2021). Relationship development with humanoid social robots: Applying interpersonal theories to human–robot interaction. *Cyberpsychology, Behavior, and Social Networking, 24*(5), 294–299. https://doi.org/10.1089/cyber.2020.0181
Fredrickson, B. L. (2001). The role of positive emotions in positive psychology: The broaden-and-build theory of positive emotions. *American Psychologist, 56*(3), 218–226. https://doi.org/10.1037/0003-066X.56.3.218
Fredrickson, B. L., & Branigan, C. (2005). Positive emotions broaden the scope of attention and thought-action repertoires. *Cognition & Emotion, 19*(3), 313–332. https://doi.org/10.1080/02699930441000238
Fritz, M. S., & MacKinnon, D. P. (2007). Required sample size to detect the mediated effect. *Psychological Science, 18*(3), 233–239. https://doi.org/10.1111/j.1467-9280.2007.01882.x
Gambino, A., Fox, J., & Ratan, R. (2020). Building a stronger CASA: Extending the computers are social actors paradigm. *Human-Machine Communication, 1*, 71–86. https://doi.org/10.30658/hmc.1.5
Gao, C. (2023). General population's psychological perceptions of COVID-19: A systematic review. *Psychology Research and Behavior Management, 16*, 4995–5009. https://doi.org/10.2147/PRBM.S440942
Gordon, A. M., & Chen, S. (2016). Do you get where i'm coming from?: Perceived understanding buffers against the negative impact of conflict on relationship satisfaction. *Journal of Personality and Social Psychology, 110*(2), 239–260. https://doi.org/10.1037/pspi0000039
Götz, F. M., Gosling, S. D., & Rentfrow, P. J. (2022). Small effects: The indispensable foundation for a cumulative psychological science. *Perspectives on Psychological Science, 17*(1), 205–215. https://doi.org/10.1177/1745691620984483
Grice, J. W. (1997). On the validity of the perceived understanding instrument. *Psychological Reports, 80*(3), 1007–1010. https://doi.org/10.2466/pr0.1997.80.3.1007
Griskevicius, V., Shiota, M. N., & Neufeld, S. L. (2010). Influence of different positive emotions on persuasion processing: A functional evolutionary approach. *Emotion, 10*(2), 190–206. https://doi.org/10.1037/a0018421
Gu, C., Zhang, Y., & Zeng, L. (2024). Exploring the mechanism of sustained consumer trust in AI chatbots after service failures: A perspective based on attribution and CASA theories. *Humanities and Social Sciences Communications, 11*(1), 1400. https://doi.org/10.1057/s41599-024-03879-5
Guerreiro, M., Barker, E., & Johnson, J. (2022). Measuring student reading comprehension performance: Considerations of accuracy, equity, and engagement by embedding comprehension items within reading passages. https://doi.org/10.7275/CH8R-TX33.
Harambam, J. (2021). Against modernist illusions: Why we need more democratic and constructivist alternatives to debunking conspiracy theories. *Journal for Cultural Research, 25*(1), 104–122. https://doi.org/10.1080/14797585.2021.1886424
Hartman, R., Moss, A. J., Jaffe, S. N., Rosenzweig, C., Litman, L., & Robinson, J. (2023). Introducing connect by CloudResearch: Advancing online participant recruitment in the digital age. *PsyArXiv*. https://doi.org/10.31234/osf.io/ksgyr
Hayes, A. F., & Little, T. D. (2022). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach* (3rd ed.). The Guilford Press.
Heidel, R. E. (2016). Causality in statistical power: Isomorphic properties of measurement, research design, effect size, and sample size. *Scientific*, 1–5. https://doi.org/10.1155/2016/8920418, 2016.
Henkel, L., Sprengholz, P., Korn, L., Betsch, C., & Böhm, R. (2022). The association between vaccination status identification and societal polarization. *Nature Human Behaviour, 7*(2), 231–239. https://doi.org/10.1038/s41562-022-01469-6
Henschke, A. (2024). *Cognitive warfare: Grey matters in contemporary political conflict* (1st ed.). Routledge. https://doi.org/10.4324/9781003126959
Heyselaar, E. (2023). The CASA theory no longer applies to desktop computers. *Scientific Reports, 13*(1), Article 19693. https://doi.org/10.1038/s41598-023-46527-9
Hornsey, M. J., Harris, E. A., & Fielding, K. S. (2018). Relationships among conspiratorial beliefs, conservatism and climate scepticism across nations. *Nature Climate Change, 8*(7), 614–620. https://doi.org/10.1038/s41558-018-0157-2
Howell, D. C. (2010). *Statistical methods for psychology* (7th ed.). Thomson Wadsworth.
Hoyle, R. H. (Ed.). (2023). *Handbook of structural equation modeling* (2nd ed.). The Guilford Press.
Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling:*

*A Multidisciplinary Journal, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118
Imhoff, R., & Lamberty, P. (2020). A bioweapon or a hoax? The link between distinct conspiracy beliefs about the coronavirus disease (COVID-19) outbreak and pandemic behavior. *Social Psychological and Personality Science, 11*(8), 1110–1118. https://doi.org/10.1177/1948550620934692
Ioku, T., & Watamura, E. (2022). Contrast of felt understanding and intergroup relations within and between communities. *Peace and Conflict: Journal of Peace Psychology, 28*(2), 245–254. https://doi.org/10.1037/pac0000605
Ioku, T., & Watamura, E. (2025). Further evidence for the role of felt understanding in intergroup relations: Japanese and Chinese relations in Japan[1]. *Japanese Psychological Research, 67*(1), 87–97. https://doi.org/10.1111/jpr.12437
Itzchakov, G., & DeMarree, K. G. (2022). Attitudes in an interpersonal context: Psychological safety as a route to attitude change. *Frontiers in Psychology, 13*. https://doi.org/10.3389/fpsyg.2022.932413
Itzchakov, G., Kluger, A. N., & Castro, D. R. (2017). I am aware of my inconsistencies but can tolerate them: The effect of high quality listening on speakers' attitude ambivalence. *Personality and Social Psychology Bulletin, 43*(1), 105–120. https://doi.org/10.1177/0146167216675339
Itzchakov, G., & Reis, H. T. (2021). Perceived responsiveness increases tolerance of attitude ambivalence and enhances intentions to behave in an open-minded manner. *Personality and Social Psychology Bulletin, 47*(3), 468–485. https://doi.org/10.1177/0146167220929218
Itzchakov, G., Reis, H. T., & Rios, K. (2024). Perceiving others as responsive lessens prejudice: The mediating roles of intellectual humility and attitude ambivalence. *Journal of Experimental Social Psychology, 110*, Article 104554. https://doi.org/10.1016/j.jesp.2023.104554
Jager, J., Putnick, D. L., & Bornstein, M. H. (2017). More than just convenient: The scientific merits of homogeneous convenience samples. *Monographs of the Society for Research in Child Development, 82*(2), 13–30. https://doi.org/10.1111/mono.12296
Jarynowski, A., Krzowski, Ł., & Maksymowicz, S. (2023). *Biological mis(dis)-information in the internet as a possible kremlin warfare* (version v3). *Zenodo*. https://doi.org/10.5281/ZENODO.7932530
Jolley, D., & Douglas, K. M. (2014). The effects of anti-vaccine conspiracy theories on vaccination intentions. *PLoS One, 9*(2), Article e89177. https://doi.org/10.1371/journal.pone.0089177
Jolley, D., & Douglas, K. M. (2017). Prevention is better than cure: Addressing anti-vaccine conspiracy theories. *Journal of Applied Social Psychology, 47*(8), 459–469. https://doi.org/10.1111/jasp.12453
Kahan, B. C., & Morris, T. P. (2012). Improper analysis of trials randomised using stratified blocks or minimisation. *Statistics in Medicine, 31*(4), 328–340. https://doi.org/10.1002/sim.4431
Kaltwasser, C. R., Taggart, P., Espejo, P. O., & Ostiguy, P. (Eds.). (2017). *The Oxford handbook of populism* (1st ed.). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198803560.001.0001.
Kirmayer, L. J. (2024a). Science and sanity: A social epistemology of misinformation, disinformation, and the limits of knowledge. *Transcultural Psychiatry, 61*(5), 795–808. https://doi.org/10.1177/13634615241296301
Kirmayer, L. J. (2024b). The fragility of truth: Social epistemology in a time of polarization and pandemic. *Transcultural Psychiatry, 61*(5), 701–713. https://doi.org/10.1177/13634615241299556
Kline, R. B. (2023). *Principles and practice of structural equation modeling* (5th ed.). Guilford.
Koller, M., & Stahel, W. A. (2017). Nonsingular subsampling for regression S estimators with categorical predictors. *Computational Statistics, 32*(2), 631–646. https://doi.org/10.1007/s00180-016-0679-x
Kordestani, A., Oghazi, P., Izmir, O., Oypan, O., & Ozer, S. (2023). Identification of the drivers of and barriers to COVID-19 vaccine intake behavior using a mixed-method design: Implications from a developing country. *Journal of Innovation & Knowledge, 8*(4), Article 100413. https://doi.org/10.1016/j.jik.2023.100413
Krispenz, A., & Bertrams, A. (2024). Understanding left-wing authoritarianism: Relations to the dark personality traits, altruism, and social justice commitment. *Current Psychology, 43*(3), 2714–2730. https://doi.org/10.1007/s12144-023-04463-x
Kumkale, G. T., & Albarracín, D. (2004). The sleeper effect in persuasion: A meta-analytic review. *Psychological Bulletin, 130*(1), 143–172. https://doi.org/10.1037/0033-2909.130.1.143
Lakoff, G. (2002). *Moral politics: How liberals and conservatives think, second edition*. University of Chicago Press. https://doi.org/10.7208/chicago/9780226471006.001.0001
Lee, I., & Hahn, S. (2024). On the relationship between mind perception and social support of chatbots. *Frontiers in Psychology, 15*, Article 1282036. https://doi.org/10.3389/fpsyg.2024.1282036
Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest, 13*(3), 106–131. https://doi.org/10.1177/1529100612451018
Lewandowsky, S., & Van Der Linden, S. (2021). Countering misinformation and fake news through inoculation and prebunking. *European Review of Social Psychology, 32*(2), 348–384. https://doi.org/10.1080/10463283.2021.1876983
Li, C.-H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods, 48*(3), 936–949. https://doi.org/10.3758/s13428-015-0619-7
Li, Q., Luximon, Y., & Zhang, J. (2023). The influence of anthropomorphic cues on patients' perceived anthropomorphism, social presence, trust building, and acceptance of health care conversational agents: Within-subject web-based experiment. *Journal of Medical Internet Research, 25*, Article e44479. https://doi.org/10.2196/44479
Lisker, M., Gottschalk, C., & Mihaljević, H. (2025). *Debunking with dialogue? Exploring AI-Generated counterspeech to challenge conspiracy theories* (no. arXiv:2504.16604). *arXiv*. https://doi.org/10.48550/arXiv.2504.16604
Livingstone, A. G. (2023). Felt understanding in intergroup relations. *Current Opinion in Psychology, 51*, Article 101587. https://doi.org/10.1016/j.copsyc.2023.101587
Livingstone, A. G., Fernández Rodríguez, L., & Rothers, A. (2020). "they just don't understand us": The role of felt understanding in intergroup relations. *Journal of Personality and Social Psychology, 119*(3), 633–656. https://doi.org/10.1037/pspi0000221
Livingstone, A. G., Windeatt, S., Nesbitt, L., Kerry, J., Barr, S. A., Ashman, L., Ayers, R., Bibby, H., Boswell, E., Brown, J., Chiu, M., Cowie, E., Doherr, E., Douglas, H., Durber, L., Ferguson, M., Ferreira, M., Fisk, I., Fleming, B., … Wu, J.-C. (2020). Do you get Us? A multi-experiment, meta-analytic test of the effect of felt understanding in intergroup relations. *Journal of Experimental Social Psychology, 91*, Article 104028. https://doi.org/10.1016/j.jesp.2020.104028
Loomba, S., De Figueiredo, A., Piatek, S. J., De Graaf, K., & Larson, H. J. (2021). Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA. *Nature Human Behaviour, 5*(3), 337–348. https://doi.org/10.1038/s41562-021-01056-1
Lun, J., Kesebir, S., & Oishi, S. (2008). On feeling understood and feeling well: The role of interdependence. *Journal of Research in Personality, 42*(6), 1623–1628. https://doi.org/10.1016/j.jrp.2008.06.009
Maechler, M., Rousseeuw, P., Croux, C., Todorov, V., & Andreas, R. (2024). Robustbase: Basic robust statistics. *Version 0.99-4-1) [Computer software]*. https://cran.r-project.org/web/packages/robustbase/index.html.
Maertens, R., Roozenbeek, J., Basol, M., & Van Der Linden, S. (2021). Long-term effectiveness of inoculation against misinformation: Three longitudinal experiments. *Journal of Experimental Psychology: Applied, 27*(1), 1–16. https://doi.org/10.1037/xap0000315
Mahjob, H., & Shakori, S. (2022). Modern cognitive warfare: From the application of cognitive science and technology in the battlefield to the arena of cognitive warfare. *Journal of Human Resource Studies, 12*(2). https://doi.org/10.22034/jhrs.2022.158895
Martin, L. R., & Petrie, K. J. (2017). Understanding the dimensions of anti-vaccination attitudes: The vaccination attitudes examination (VAX) scale. *Annals of Behavioral Medicine, 51*(5), 652–660. https://doi.org/10.1007/s12160-017-9888-y
McCright, A. M., Marquart-Pyatt, S. T., Shwom, R. L., Brechin, S. R., & Allen, S. (2016). Ideology, capitalism, and climate: Explaining public views about climate change in the United States. *Energy Research & Social Science, 21*, 180–189. https://doi.org/10.1016/j.erss.2016.08.003
McGuire, W. J. (1961). Resistance to persuasion conferred by active and passive prior refutation of the same and alternative counterarguments. *Journal of Abnormal and Social Psychology, 63*(2), 326–332. https://doi.org/10.1037/h0048344
McGuire, J., De Cremer, D., Hesselbarth, Y., De Schutter, L., Mai, K. M., & Van Hiel, A. (2023). The reputational and ethical consequences of deceptive chatbot use. *Scientific Reports, 13*(1), Article 16246. https://doi.org/10.1038/s41598-023-41692-3
Minson, J. A., & Chen, F. S. (2022). Receptiveness to opposing views: Conceptualization and integrative review. *Personality and Social Psychology Review, 26*(2), 93–111. https://doi.org/10.1177/10888683211061037
Minson, J. A., Chen, F. S., & Tinsley, C. H. (2020). Why won't you listen to me? Measuring receptiveness to opposing views. *Management Science, 66*(7), 3069–3094. https://doi.org/10.1287/mnsc.2019.3362
Morelli, S. A., Torre, J. B., & Eisenberger, N. I. (2014). The neural bases of feeling understood and not understood. *Social Cognitive and Affective Neuroscience, 9*(12), 1890–1896. https://doi.org/10.1093/scan/nst191
Muszyński, M. (2023). Attention checks and how to use them: Review and practical recommendations. *Ask: Research and Methods, 32*(1), 3–38. https://doi.org/10.18061/ask.v32i1.0001
Muthén, L. K., & Muthén, B. O. (2002). How to use a monte carlo study to decide on sample size and determine power. *Structural Equation Modeling: A Multidisciplinary Journal, 9*(4), 599–620. https://doi.org/10.1207/S15328007SEM0904_8
Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues, 56*(1), 81–103. https://doi.org/10.1111/0022-4537.00153
Nwokolo, S. C. (2025). Climate hoax: The shift from scientific discourse to speculative rhetoric in climate change conversations. *Research: Ideas for Today's Investors, 2*(2), Article 100322. https://doi.org/10.1016/j.nexres.2025.100322
Oishi, S., Krochik, M., & Akimoto, S. (2010). Felt understanding as a bridge between close relationships and subjective well-being: Antecedents and consequences across individuals and cultures: Felt understanding as a bridge. *Social and Personality Psychology Compass, 4*(6), 403–416. https://doi.org/10.1111/j.1751-9004.2010.00264.x
Orinx, K., & Struye de Swielande, T. (2022). China and cognitive warfare: Why is the west losing? In B. Claverie, B. Prébot, N. Beuchler, & F. du Cluzel (Eds.), *Cognitive warfare: The future of cognitive dominance*. NATO Collaboration Support Office, 8, 1–6.
Palecek, M., & Hampel, V. (2024). Conspiracy theories and anxiety in culture: Why is threat-related misinformation an evolved product of our ability to mobilize sources in the face of Un-represented threat? *Philosophy of the Social Sciences, 54*(2), 99–132. https://doi.org/10.1177/00483931231210335
Parezanović, M., & Proroković, D. (2024). Psychological and propaganda operations as a form of hybrid warfare. *Nauka, 29*(1), 43–55. https://doi.org/10.5937/nabepo29-45316. Bezbednost, Policija.

Peer, E., Rothschild, D., Gordon, A., Evernden, Z., & Damer, E. (2021). Data quality of platforms and panels for online behavioral research. *Behavior Research Methods, 54* (4), 1643–1662. https://doi.org/10.3758/s13428-021-01694-3

Petty, R. E., & Briñol, P. (2015). Emotion and persuasion: Cognitive and meta-cognitive processes impact attitudes. *Cognition & Emotion, 29*(1), 1–26. https://doi.org/10.1080/02699931.2014.967183

Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. *Advances in Experimental Social Psychology, 19*, 123–205. https://doi.org/10.1016/S0065-2601(08)60214-2. Elsevier.

Petty, R. E., Haugtvedt, C. P., & Smith, S. M. (1995). Elaboration as a determinant of attitude strength: Creating attitudes that are persistent, resistant, and predictive of behavior. In *Attitude strength: Antecedents and consequences* (pp. 93–130). Lawrence Erlbaum Associates, Inc.

Petty, R. E., Schumann, D. W., Richman, S. A., & Strathman, A. J. (1993). Positive mood and persuasion: Different roles for affect under high- and low-elaboration conditions. *Journal of Personality and Social Psychology, 64*(1), 5–20. https://doi.org/10.1037/0022-3514.64.1.5

Pigden, C. (2024). How to make conspiracy theory research intellectually respectable (and what it might be like if it were). *Inquiry*, 1–25. https://doi.org/10.1080/0020174X.2024.2375780

Pocheptsov, G. (2018). Cognitive attacks in Russian hybrid warfare. *Information and security : an international journal, 41*, 37–43. https://doi.org/10.11610/isij.4103

Popper, K. R. (1984). *Logik der Forschung (8., weiter verb. und verm.* Aufl: Mohr.

Primbs, M. A., Pennington, C. R., Lakens, D., Silan, M. A. A., Lieck, D. S. N., Forscher, P. S., Buchanan, E. M., & Westwood, S. J. (2023). Are Small Effects the Indispensable Foundation for a Cumulative Psychological Science? A Reply to Götz et al. *Perspectives on Psychological Science, 18*(2), 508–512. https://doi.org/10.1177/17456916221100420, 2022.

Rao Hill, S., & Troshani, I. (2024). Chatbot anthropomorphism, social presence, uncanniness and brand attitude effects. *Journal of Computer Information Systems*, 1–17. https://doi.org/10.1080/08874417.2024.2423187

Reeves, B., & Nass, C. I. (1996). *The media equation: How people treat computers, television, and new media like real people and places* (1. paperback ed.). CSLI Publ [reprint.]).

Reis, H. T., Carmichael, C. L., Rogge, R. D., Maniaci, M. R., & Crasta, D. (2017). Perceived partner responsiveness scale (PPRS). In D. L. Worthington, & G. D. Bodie (Eds.), *The sourcebook of listening research* (1st ed., pp. 516–521). Wiley. https://doi.org/10.1002/9781119102991.ch57.

Reis, H. T., & Gable, S. L. (2015). Responsiveness. *Current Opinion in Psychology, 1*, 67–71. https://doi.org/10.1016/j.copsyc.2015.01.001

Reschke, B., Minson, J. A., Bowles, H. R., De Vaan, M., & Srivastava, S. B. (2020). Mutual receptiveness to opposing views bridges ideological divides in network formation. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3703958

Rhemtulla, M., Brosseau-Liard, P.É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods, 17*(3), 354–373. https://doi.org/10.1037/a0029315

Rheu, M., Dai, Y.(N.), Meng, J., & Peng, W. (2024). When a chatbot disappoints you: Expectancy violation in human-chatbot interaction in a social support context. *Communication Research, 51*(7), 782–814. https://doi.org/10.1177/00936502231221669

Roos, C. A., Postmes, T., & Koudenburg, N. (2023). Feeling heard: Operationalizing a key concept for social relations. *PLoS One, 18*(11), Article e0292865. https://doi.org/10.1371/journal.pone.0292865

Roozenbeek, J., Van Der Linden, S., Goldberg, B., Rathje, S., & Lewandowsky, S. (2022). Psychological inoculation improves resilience against misinformation on social media. *Science Advances, 8*(34). https://doi.org/10.1126/sciadv.abo6254. eabo6254.

Roozenbeek, J., Van Der Linden, S., & Nygren, T. (2020). Prebunking interventions based on the psychological theory of "inoculation" can reduce susceptibility to misinformation across cultures. *Harvard Kennedy School Misinformation Review*. https://doi.org/10.37016//mr-2020-008

Rosseel, Y., Jorgensen, T. D., Wilde, L. D., Oberski, D., Byrnes, J., Vanbrabant, L., Savalei, V., Merkle, E., Hallquist, M., Rhemtulla, M., Katsikatsou, M., Barendse, M., Rockwood, N., Scharf, F., Du, H., Jamil, H., & Classe, F. (2024). Lavaan: Latent variable analysis [Computer software] Version 0.6-19. https://cran.r-project.org/web/packages/lavaan/index.html.

Rucker, D. D., Preacher, K. J., Tormala, Z. L., & Petty, R. E. (2011). Mediation analysis in social psychology: Current practices and new recommendations. *Social and Personality Psychology Compass, 5*(6), 359–371. https://doi.org/10.1111/j.1751-9004.2011.00355.x

Rutjens, B. T., & Hornsey, M. (2024). The psychology of science rejection. https://doi.org/10.31234/osf.io/cz4nf.

Sass, D. A., Schmitt, T. A., & Marsh, H. W. (2014). Evaluating model fit with ordered categorical data within a measurement invariance framework: A comparison of estimators. *Structural Equation Modeling: A Multidisciplinary Journal, 21*(2), 167–180. https://doi.org/10.1080/10705511.2014.882658

Savalei, V. (2021). Improving fit indices in structural equation modeling with categorical data. *Multivariate Behavioral Research, 56*(3), 390–407. https://doi.org/10.1080/00273171.2020.1717922

Schlegel, K., Sommer, N. R., & Mortillaro, M. (2025). Large language models are proficient in solving and creating emotional intelligence tests. *Communications Psychology, 3*(1), 80. https://doi.org/10.1038/s44271-025-00258-x

Schrodt, P. (2003). Student perceptions of instructor verbal aggressiveness: The influence of student verbal aggressiveness and self-esteem. *Communication Research Reports, 20* (3), 240–250. https://doi.org/10.1080/08824090309388822

Schrodt, P., & Finn, A. N. (2011). Students' perceived understanding: An alternative measure and its associations with perceived teacher confirmation, verbal aggressiveness, and credibility. *Communication Education, 60*(2), 231–254. https://doi.org/10.1080/03634523.2010.535007

Schwarz, N., & Bless, H. (1991). Happy and mindless, but sad and smart? The impact of affective states on analytic reasoning. In J. P. Forgas (Ed.), *Emotion and social judgments* (1st ed., pp. 55–71). Garland Science. https://doi.org/10.4324/9781003058731-4.

Seehausen, M., Kazzer, P., Bajbouj, M., Heekeren, H. R., Jacobs, A. M., Klann-Delius, G., Menninghaus, W., & Prehn, K. (2014). Talking about social conflict in the MRI scanner: Neural correlates of being empathized with. *NeuroImage, 84*, 951–961. https://doi.org/10.1016/j.neuroimage.2013.09.056

Shamon, H., & Berning, C. C. (2020). Attention check items and instructions in online surveys: Boon or bane for data quality? *Survey Research Methods, 55–77*. https://doi.org/10.18148/SRM/2020.V14I1.7374

Şimşek, C. (2024). Voices unheard: How feelings of inefficacy fuel populism. *Comparative European Politics, 22*(5), 662–683. https://doi.org/10.1057/s41295-024-00378-4

Swann, J. W. B. (1990). To be adored or to be known? The interplay of self-enhancement and self-verification. *Handbook of motivation and cognition: Foundations of social behavior, 2*, 408–448. The Guilford Press.

Tam, L., & Kim, S. (2023). Understanding conspiratorial thinking (CT) within public relations research: Dynamics of organization-public relationship quality, CT, and negative megaphoning. *Public Relations Review, 49*(4), Article 102354. https://doi.org/10.1016/j.pubrev.2023.102354

Tashev, B., Purcell, M., & McLaughlin, B. (2019). Russia's information warfare: Exploring the cognitive dimension. *MCU Journal, 10*(2), 129–147. https://doi.org/10.21140/mcuj.2019100208

Traberg, C. S., Roozenbeek, J., & Van Der Linden, S. (2022). Psychological inoculation against misinformation: Current evidence and future directions. *The Annals of the American Academy of Political and Social Science, 700*(1), 136–151. https://doi.org/10.1177/00027162221087936

Tutz, G. (2022). Ordinal regression: A review and a taxonomy of models. *WIREs Computational Statistics, 14*(2), Article e1545. https://doi.org/10.1002/wics.1545

Tyson, A., Funk, C., & Kennedy, B. (2023). *What the data says about americans' views of climate change*. Pew Research Center. https://www.pewresearch.org/short-reads/2023/08/09/what-the-data-says-about-americans-views-of-climate-change/.

U.S. Senate Select Committee on Intelligence. (2017). *Report of the select Committee on intelligence on Russian active measures campaigns and interference in the 2016 U.S. election, volume 2: Russia's use of social media with additional views*. U.S. Government Publishing Office. https://web.archive.org/web/20191012052520/https://www.intelligence.senate.gov/sites/default/files/documents/Report_Volume2.pdf.

Većkalov, B., Geiger, S. J., Bartoš, F., White, M. P., Rutjens, B. T., Van Harreveld, F., Stablum, F., Akın, B., Aldoh, A., Bai, J., Berglund, F., Bratina Zimic, A., Broyles, M., Catania, A., Chen, A., Chorzępa, M., Farahat, E., Götz, J., Hoter-Ishay, B., … Van Der Linden, S. (2024). A 27-country test of communicating the scientific consensus on climate change. *Nature Human Behaviour, 8*(10), 1892–1905. https://doi.org/10.1038/s41562-024-01928-2

Velez, Y. R., & Liu, P. (2024). Confronting Core issues: A critical assessment of attitude polarization using tailored experiments. *American Political Science Review*, 1–18. https://doi.org/10.1017/S0003055424000819

Walter, N., & Tukachinsky, R. (2020). A meta-analytic examination of the continued influence of misinformation in the face of correction: How powerful is It, why does it happen, and how to stop it? *Communication Research, 47*(2), 155–177. https://doi.org/10.1177/0093650219854600

Webb, M. A., & Tangney, J. P. (2024). Too good to be true: Bots and bad data from mechanical turk. *Perspectives on Psychological Science, 19*(6), 887–890. https://doi.org/10.1177/17456916221120027

Whittaker, T. A., & Schumacker, R. E. (2022). *A beginner's guide to structural equation modeling* (5th ed.). Routledge.

Wolsko, C., Ariceaga, H., & Seiden, J. (2016). Red, white, and blue enough to be green: Effects of moral framing on climate change attitudes and conservation behaviors. *Journal of Experimental Social Psychology, 65*, 7–19. https://doi.org/10.1016/j.jesp.2016.02.005

Wood, T., & Porter, E. (2016). The elusive backfire effect: Mass attitudes' steadfast factual adherence. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2819073

Worth, L. T., & Mackie, D. M. (1987). Cognitive mediation of positive affect in persuasion. *Social Cognition, 5*(1), 76–94. https://doi.org/10.1521/soco.1987.5.1.76

Wygnańska, J. (2023). The experience of conversation and relation with a well-being chabot: Between proximity and remoteness. *Qualitative Sociology Review, 19*(4), 92–120. https://doi.org/10.18778/1733-8077.19.4.05

Yeomans, M., Minson, J. A., 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

Yin, Y., Jia, N., & Wakslak, C. J. (2024). AI can help people feel heard, but an AI label diminishes this impact. *Proceedings of the National Academy of Sciences, 121*(14), Article e2319112121. https://doi.org/10.1073/pnas.2319112121

Zembylas, M. (2023). Moving beyond debunking conspiracy theories from a narrow epistemic lens: Ethical and political implications for education. *Pedagogy, Culture & Society, 31*(4), 741–756. https://doi.org/10.1080/14681366.2021.1948911

Zhang, B., & Gearhart, S. (2020). Collecting Online Survey data: A comparison of data quality among a commercial Panel & MTurk. *Survey Practice, 13*(1), 1–10. https://doi.org/10.29115/SP-2020-0015

Zilinsky, J., Theocharis, Y., Pradel, F., Tulin, M., De Vreese, C., Aalberg, T., Cardenal, A. S., Corbu, N., Esser, F., Gehle, L., Halagiera, D., Hameleers, M., Hopmann, D. N., Koc-Michalska, K., Matthes, J., Schemer, C., Štětka, V.,

Strömbäck, J., Terren, L., … Zoizner, A. (2024). Justifying an invasion: When is disinformation successful? *Political Communication*, 1–22. https://doi.org/10.1080/10584609.2024.2352483