



ARTICLE



<https://doi.org/10.1057/s41599-024-03416-4>

OPEN

# Identifying the underlying psychological constructs from self-expressed anti-vaccination argumentation

Dawn Holford <sup>1✉</sup>, Ezequiel Lopez-Lopez<sup>2</sup>, Angelo Fasce<sup>3</sup>, Linda C. Karlsson<sup>4</sup> & Stephan Lewandowsky <sup>1,5,6</sup>

People's negative attitudes to vaccines can be motivated by psychological factors—such as fears, ideological beliefs, and cognitive patterns—known as 'attitude roots'. This study had two primary objectives: (1) to identify which of 11 known attitude roots are featured in individuals' self-expressed reasons for negative vaccine attitudes (i.e., a linguistic analysis); (2) to explore how attitude roots present in self-expressed texts are linked to specific psychological measures. To achieve Objective 1, our study collected data from December 2022 to January 2023 from 556 participants from the US, who wrote texts to explain the reasons for their negative vaccine attitudes. The texts encompassed 2327 conceptually independent units of anti-vaccination argumentation, that were each coded for its attitude root(s) by at least two psychological experts. By allowing participants to spontaneously express their attitudes in their own words, we were able to observe how this differed from what participants reported to endorse when presented with a list of arguments. We found that there were four groups of attitude roots based on linguistic similarity in self-expression. In addition, latent class analysis of participants' coded texts identified three distinct groups of participants that were characterised by their tendency to express combinations of arguments related to (1) fears, (2) anti-scientific conceptions, and (3) politicised perspectives. To achieve Objective 2, we collected participants' responses to 11 validated measures of psychological constructs expected to underlie the respective 11 attitude roots, and used a correlational design to investigate how participants' self-expressed attitude roots were linked to these measures. Logistic regressions showed that an expected psychological construct was the strongest, and significant, predictor for expression of three out of the four attitude root groups. We discuss the implications of these findings for health communicators and practitioners.

<sup>1</sup>School of Psychological Science, University of Bristol, Bristol, England. <sup>2</sup>Max Planck Institute for Human Development, Berlin, Germany. <sup>3</sup>Faculty of Medicine, University of Coimbra, Coimbra, Portugal. <sup>4</sup>Institute of Clinical Medicine, University of Turku, Turku, Finland. <sup>5</sup>Department of Psychology, University of Potsdam, Potsdam, Germany. <sup>6</sup>School of Psychological Sciences, University of Western Australia, Crawley, WA, Australia.  
✉email: [dawn.holford@bristol.ac.uk](mailto:dawn.holford@bristol.ac.uk)

## Introduction

Despite decades of evidence that vaccination is one of the most effective and life-saving public health interventions (Greenwood, 2014), argumentation against vaccination persists, often fuelled by misinformation. The propagation of such arguments has tangible costs for society, as it may lower vaccine uptake, thereby threatening public health campaigns and the control of vaccine-preventable diseases (Broniatowski et al., 2018; Council of Canadian Academies, 2023). A recent modelling study from Canada has estimated the cost of COVID-19 misinformation to be 2800 preventable deaths and an additional burden of \$299 million in preventable medical costs (Council of Canadian Academies, 2023). This paper aims to help identify what motivates people to believe and express arguments against vaccination. This has been a critical question for psychological researchers, as it is firmly established that simply providing factual, scientific information about the benefits of vaccination is insufficient to overturn people's anti-vaccination beliefs (Ecker et al., 2022; Karlsson et al., 2024; Ruggeri et al., 2024; Soveri et al., 2023).

One theory for why people persist with misinformed beliefs in the face of scientific evidence is the 'attitude root theory' (Hornsey, 2020). According to this theory, individuals' dismissal of evidence is motivated by an underlying psychological factor, known as an 'attitude root'. Attitude roots cover a range of psychological constructs, such as fears, ideologies, and worldviews. Indeed, studies have shown that individuals assess information in the context of their own worldviews, values, and beliefs, which can result in a motivation to favour information confirming their beliefs and dismiss that which contradicts those beliefs (Lewandowsky and Oberauer, 2016). Attitude roots may be identified in a top-down way (e.g., by looking at psychological constructs that predict vaccine attitudes; Hornsey et al., 2018) or a bottom-up way (e.g., by examining argumentation against vaccination; Fasce et al., 2023). A systematic literature review that integrated multiple previous typologies of anti-vaccination argumentation identified 11 candidate attitude roots motivating such argumentation, described in Table 1 (Fasce et al., 2023).

According to attitude root theory, if an individual scores high on a psychological factor (i.e., attitude root), they would be likely to express an argument that linguistically reflects that attitude root. For example, a tendency to experience anxiety or fear over medical procedures may motivate individuals to express doubts about the safety record of vaccines, whereas individuals high on 'trait reactance' (the tendency to experience oppositional emotions and cognitions towards perceived threat to one's autonomy) may be motivated to argue against vaccine recommendations on the grounds that these restrict freedom of choice. One would also expect the reverse to be observed: if an individual expresses argumentation that linguistically reflects an attitude root, they would likely score highly on measures of a psychological construct related to that root. Previous research has linked people's style of linguistic expression to a variety of individual attributes. For example, people with similar personality traits tend to write or choose to read text that is linguistically similar, to the extent that researchers can predict these traits from text samples with a remarkable degree of accuracy (Boyd and Pennebaker, 2017; Noecker et al., 2013; Simchon et al., 2023). Given that attitude roots, like personality, are assumed to be relatively stable characteristics of a person, it seems likely that they, too, find expression in people's produced text explaining their vaccine attitudes. In support of this possibility, at least one relevant psychological construct (namely people's 'epistemic' beliefs about the nature of truth and honesty) has been shown to be associated with the way individuals expressed themselves on Twitter (Lasser et al., 2023).

The ability to identify an individual's attitude roots when interacting with them is important because research has shown that vaccine communication is more effective if it is aligned with individuals' attitude roots. To illustrate, attitude root theory would expect that an individual who endorses natural approaches to health is presented with scientific evidence about the safety of a new mRNA vaccine might reject this information because it does

**Table 1** Coding framework for attitude roots in texts produced by participants.

Attitude root	Description
Conspiracist ideation	The tendency to believe in conspiracy theories. Arguments in this category reflect suspicion about (usually nefarious) plot(s) and the presence of malevolent actors.
Distrust	General mistrust in various sources or communicators. Arguments in this category reflect a (sometimes vague) sense of suspicion or uncertainty in the source of information. Although conspiracy implies distrust, distrust need not indicate conspiracy is present.
Unwarranted beliefs	This category captures a variety of beliefs that are not backed up by science, misrepresent scientific facts, or rely on pseudoscientific conceptions. Unwarranted belief arguments were considered as differentiated from distorted risk perception in that they will talk about the alternative being 'better' rather than 'as good as'.
Worldview and politics	Arguments in this category typically reference an individual's ideological view on how society is, or should be, organised, and a sense that vaccination goes against this view.
Religious concerns	Arguments in this category encompass a range of religious and/or spiritual beliefs or norms that vaccination is believed to contravene.
Moral concerns	This category includes arguments related to an individual's moral stance, often issues where vaccine is seen to promote the immoral position.
Fear and phobias	This category captures the different fears that lead individuals to reject vaccines. These are usually related to worries about suffering consequences from having a vaccine.
Distorted risk perception	This category reflects a skewed risk-benefit calculation, where the argument portrays the risk of the disease is perceived to be lower and the risk of the vaccine to be higher than it actually is.
Perceived self-interest	This category captures a prioritisation of an individual's own needs relative to others.
Epistemic relativism	This category includes arguments that claim concepts like truth, facts, logic, and reasoning are subjective and should not take precedence in vaccine-related decision-making. Coding for epistemic relativism also considered the primacy of individuals' personal judgement and experiences when evaluating information.
Reactance	This category captures a tendency to defend one's autonomy when their freedom is perceived to be restricted (e.g., because they are asked to get vaccinated).

The coding framework was based on Fasce et al. (2023)'s taxonomy of attitude roots. Coders also referred to themes found under each of the roots, which are documented in that paper.

**Table 2 Mean (SD), internal consistency, and source of psychological construct variables.**

Psychological trait	Internal consistency	Mean/maximum	SD	Source of scale
Conspiracist mentality	0.77	0.73/1	0.18	Bruder et al. (2013)
Distrust	0.82	2.72/5	0.94	Yamagishi (1988)
Pseudoscientific beliefs	0.77	3.06/5	0.99	Fasce et al. (2021)
Populist attitudes	0.59	4.03/5	0.73	Akkerman et al. (2013)
Centrality of religion	0.89	3.31/5	1.28	Huber and Huber (2012)
Moral absolutism	0.72	2.32/4	0.77	Peterson et al. (2009)
General medical anxiety	0.52	2.72/5	0.94	van Balen and Verdurmen (1999)
Risk distortion	0.53 <sup>a</sup>	22.84/100	30.05	Betsch et al. (2018); Schmid et al. (2017)
Self-interest	0.82	4.59/7	1.39	Gerbasi and Prentice (2013)
Epistemic belief (Truth is political)	0.86	3.17/5	1.28	Garrett and Weeks (2017)
Reactance	0.63	3.15/5	0.83	Hong and Faedda (1996); Hornsey et al. (2018)

Mean and SD columns show unstandardised scores, with the maximum score indicated; see Supplementary File 2, Table S11 for details of all response scales. Internal consistency is measured as Cronbach’s  $\alpha$  for all items except risk distortion.

<sup>a</sup>Risk distortion was measured as the difference between perceived risk of two vaccines and the perceived risk of two diseases. As there were only two items for risk distortion, we present here the correlation (Pearson’s  $r$ ) between these items rather than  $\alpha$ .

not align with their beliefs that only natural products are safe. However, according to the theory, presenting the scientific information in a tailored manner that underscores the fact that vaccines stimulate the body into promoting natural defences would likely make the individual more willing to accept the vaccine (Holford et al., 2024; Hornsey, 2020). A growing body of evidence supports the effectiveness of tailoring communication to address people’s attitude roots (e.g., Attwell and Freeman, 2015; Chu et al., 2021; Holford et al., 2024). Unvaccinated individuals reported greater trust in medical professionals discussing vaccines when those professionals first expressed their shared religiosity (Chu et al., 2021), or affirmed the validity of the individual’s attitude root (Holford et al., 2024), compared to when they did not. An Australian community campaign that positioned vaccination as congruent with the values of alternative lifestyle communities—engaging with their ideology and identity—also reported positive impact on vaccination views for three quarters of their targeted population (Attwell and Freeman, 2015). For communicators, such as health practitioners, to respond in an attitude root congruent manner, they must first be able to identify the relevant attitude root. It is therefore useful for them to know how best to approach this identification problem, whether by assessing individuals’ agreement to certain anti-vaccination arguments, or by analysing the content of what an individual says.

The link between an expression of certain attitude roots in argumentation and the underlying psychological construct posited to be motivating it is not always clear. To begin with, even different attitude roots that are conceptually distinct exhibit some overlap with other attitude roots (Fasce et al., 2023). For example, the attitude root of ‘conspiracist ideation’ is distinct from ‘distrust’ in that one may be distrustful without tending to believe that there is a nefarious plot; however, conspiracist ideation necessitates a large amount of interpersonal or intergroup distrust as a psychological antecedent. Individual arguments can also encompass multiple attitude roots. For example, the pervasive myth that vaccines cause implausible ‘injuries’ such as autism is rooted both in fear as well as unwarranted beliefs about cause and effect.

Individuals can also score high on multiple psychological construct measures representing different attitude roots. A tendency to believe in conspiracy theories has been found to be significantly associated with hierarchical worldviews, intuitive belief patterns, religiosity, and fear and anxieties (Goreis and Voracek, 2019), as well as reactance (Gerace et al., 2022; Hornsey et al., 2018), which in turn is also associated with distrust in government (Gerace et al., 2022; Soveri et al., 2020) and beliefs in

unevidenced health practices such as complementary and alternative medicine (Soveri et al., 2020). A study of psychological constructs corresponding to 11 attitude roots found that conspiracist ideation, distrust, reactance, populist worldviews, and beliefs in alternative epistemology were all significantly correlated (Holford et al., 2023).

There are thus two key challenges to mapping self-expressions of attitude roots (i.e., an observed behaviour) to the psychological construct (i.e., an individual’s motivation for that behaviour): First, self-expressed argumentation can be conceptually mapped to multiple attitude roots (Fasce et al., 2023). Second, many psychological constructs are interrelated (Holford et al., 2023). Previous investigations have been constrained in their ability to tackle these challenges because their data did not include individual difference measures *and* argumentation expressed first-hand by individuals. The anti-vaccination text analysed in Fasce et al. (2023)’s study was not linked to psychological traits, whereas the psychological constructs analysed in Holford et al. (2023)’s study were linked to endorsements of a pre-determined set of anti-vaccination arguments rather than self-expression of arguments.

The present study resolved the problems of previous work by collecting self-expressed text produced by individuals explaining their attitudes towards vaccination so that these could be associated with measures of the relevant psychological constructs. In other words, participants had to express their argumentation in their own words rather than simply rate arguments presented to them. The first study objective was to identify which of the 11 attitude roots in Fasce et al.’s (2023) taxonomy were featured linguistically in participants’ self-expressed text. The second objective was to explore associations between linguistic expressions of attitude roots (in their self-expressed text) and specific psychological constructs. Our pre-registered hypothesis was that self-expression of attitude roots in free text would be predicted by participants’ scores on psychological construct measures (Table 2) reflecting each attitude root in Table 1. In addition, we explored similarities in self-expression of different attitude roots through a machine learning analysis of the linguistic features of argumentation, as well as a latent class analysis of individuals based on their choice of different arguments to express in their overall text explanations.

**Methods**

The study was approved by the University of Bristol’s Research Ethics Committee (reference: 10708). Informed consent was

**Table 3 Reliability of coding for attitude roots.**

Attitude root	Round 1	Round 2	Final coder reliability	Root frequency
Conspiracist ideation	0.36	0.71	0.67	0.037
Distrust	0.65	0.71	0.72	0.439
Unwarranted beliefs	0.54	0.58	0.65	0.107
Worldview & politics	0.11	0.51	0.64	0.032
Religious concerns	0.00	0.00	0.79	0.004
Moral concerns	0.44	0.59	0.27	0.016
Fear & phobias	0.65	0.75	0.71	0.195
Distorted risk perception	0.74	0.78	0.78	0.125
Perceived self-interest	0.00	0.13	0.70	0.004
Epistemic relativism	0.57	0.47	0.60	0.144
Reactance	0.88	0.75	0.83	0.135
Overall	0.45	0.54	0.67	-
Overall (weighted) <sup>a</sup>	-	-	0.71	-

Given the prominent class imbalance of the dataset, the fuzziness of the classes often conceptually overlapping in a multi-label scenario, and the nuance of their definitions, the overall coder reliability 0.67 is already a substantial degree of agreement across raters.

<sup>a</sup>Considering the high imbalance of the roots in the dataset, we also calculated a reliability score weighted by frequency of occurrence of each root.

obtained from all participants prior to their participation. Data collection was carried out between 8 December 2022 and 9 January 2023. The study methods and analyses were pre-registered prior to data collection. The pre-registration, materials, data, and analysis scripts for results reported in this manuscript are available on the Open Science Framework: <https://osf.io/yfks3/>.

**Participants.** We recruited 612 US participants to complete the study via Prolific, inviting only respondents who had previously indicated they were ‘negative’ or ‘neutral’ (or ‘not stated’) towards the COVID-19 vaccine. Participants were paid at a rate of £9/h to complete the ~15-min study and offered bonus payments if they provided highly detailed responses to an open-ended question. The bonus rate was £0.25 and £0.50 on top of the participation fee for responses exceeding 100 and 200 words, respectively.

Following our pre-registered exclusion criteria, we excluded participants who only expressed positive arguments for vaccination ( $n = 40$ ) and therefore provided no anti-vaccination arguments for the analysis. Three participants were excluded on initial screening because one provided an off-topic response (about going on vacation), one provided a response that was verifiably copied and pasted from official health sources, and the last did not complete all the psychometric questionnaire measures. Another 13 participants were excluded during the data coding process because all coders agreed that there was no substantial argumentative content provided in their text that could be reasonably coded. This left us with a final analytical sample of 556, which was slightly under our targeted pre-registered sample size ( $n = 570$ ).

Our sample included 54% women (45% men, 1% preferred not to state), 81% White (7% Black, 6% Mixed or other ethnicities, 6% preferred not to state), had an age range of 18–76 years ( $M = 40.37$ ,  $SD = 13.12$ ), and leaned politically right based on their affiliations (44% identified as Republican, 49% Independent/other parties, 8% Democrat), although self-reported political leaning was normally distributed (skewness =  $-0.24$ , kurtosis =  $-0.19$ ).<sup>1</sup>

**Materials and procedure.** Participants completed the study anonymously online using Qualtrics. They first indicated their attitude towards vaccination in general on a 7-point scale (1 = strongly negative, 7 = strongly positive). Although we were only able to recruit participants based on their specific COVID-19 vaccination attitudes, we expected that within this recruitment target, the majority of participants would also report negative

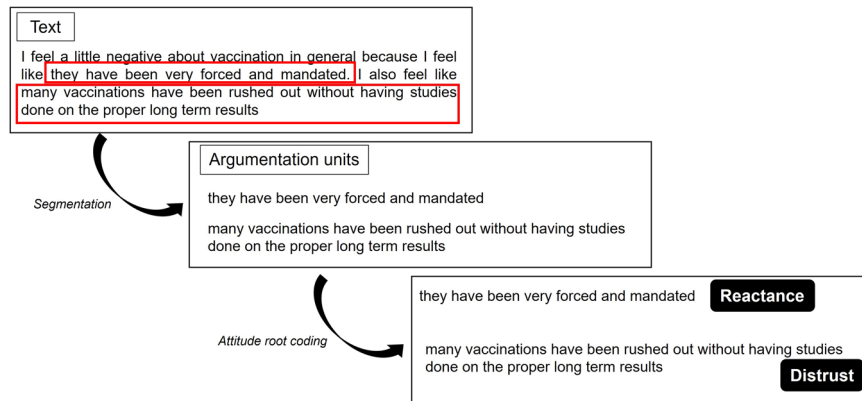
general vaccination attitudes (i.e., scoring  $< 4$ ), and this was indeed the case (67%).

On the next screen, participants then responded to a text elicitation question where they were asked to write their reasons for their vaccination attitude. Participants were informed they would receive a bonus payment for writing more detailed answers to this question. A timer on the page also prevented them from moving forward until at least one minute had elapsed. On average, participants spent a median of 4 minutes writing and produced a mean of 115 words ( $SD = 79$ ) for the text response.

Participants next completed a questionnaire that included 11 psychological construct measures (see Table 3) and ratings of agreement and familiarity with 11 prototypical anti-vaccination arguments that represented each of the attitude roots.<sup>2</sup> To constrain the overall length of the questionnaire, we limited the length of the measurement scales for the psychological constructs to a maximum of three items each, based on the items with the highest factor loadings for the construct as indicated in previous research. Reliability of all scales was satisfactory ( $> 0.60$ ) except for populist attitudes ( $\alpha = 0.56$ ; dropping one item improved this to 0.73) and general medical anxiety ( $\alpha = 0.52$ ). Given the unidimensional nature of the constructs (observed in previous studies and also in this sample through a series of factor analyses), the items of each scale were averaged to create a mean score per participant for each construct. Each of the construct measures (and the argument rating measures) was presented as a block, with the order of presentation of the blocks randomised for each participant. The exact wording of all questionnaire items and their response options can be found in Supplementary Materials.

Finally, participants completed socio-demographic information before being debriefed.

**Coding of attitude roots in expressed text.** On reviewing the texts produced by the participants, we observed that there was great variety in the structure of argumentation produced, from a couple of words to full sentences or arguments spanning multiple run-on sentences. This required us to first segment texts into individual units of anti-vaccination argumentation before coding them for attitude roots (see Fig. 1). This segmentation was completed by three experts in anti-vaccination argumentation (DH, AF, LK), who independently segmented participants’ texts into spans representing individual units of anti-vaccination argumentation, according to an initial expert-informed definition of anti-vaccination argument based on Fasce et al. (2023).<sup>3</sup>



**Fig. 1** Two-stage process of segmenting texts followed by coding of attitude roots for the segmented argumentation units.

Guidelines were determined to preserve the smallest possible stand-alone unit of argumentation that retained sufficient interpretable meaning. Coders first completed 50 texts and then met to discuss these and revise the segmentation rules for maximum clarity. A second round was then completed for another 50 texts. At this stage, critical aspects of the guidelines were clarified, and coders’ agreement was satisfactory (average Jaccard index of 0.61 for the first round and 0.72 for the second round)<sup>4</sup> and the three coders proceeded to segment all the texts independently. When all segmentation was complete, a majority rule (i.e., at least two coders agreed) was used to determine the final segmented arguments. If coders were not aligned, a decision was made through discussion with all coders and an additional arbiter (EL) familiar with the work. Argument borders (i.e., their start and end tokens) were checked again by the independent arbiter (EL) after the agreement. The final segmented arguments had an average Jaccard index of 0.67, which is considered a high level of similarity in how coders segmented the texts into arguments (Verma and Aggarwal, 2020). From the 556 texts, 2327 units of argumentation were extracted, with an average 2.97 arguments per text.

The three coders then proceeded to code each individual argument segment for the attitude roots. The coding framework was based on Fasce et al. (2023)’s anti-vaccination argument taxonomy (Table 1). All coders first independently coded the same 200 arguments and met to discuss discrepancies among these and revise the coding framework to achieve greater consistency and replicability. A second set of 200 arguments was then coded as another check, where upon satisfactory reliability among the three coders was achieved. The coders then divided all arguments among them such that every argument was coded by two of the coders, to allow for coding reliability to be checked (reported in Table 3). Disagreements among coders were resolved through discussion.

Once attitude roots had been coded for all the argument segments, we assigned these codes to the original texts, such that each text was coded for the attitude roots expressed within its composite arguments.

**Results**

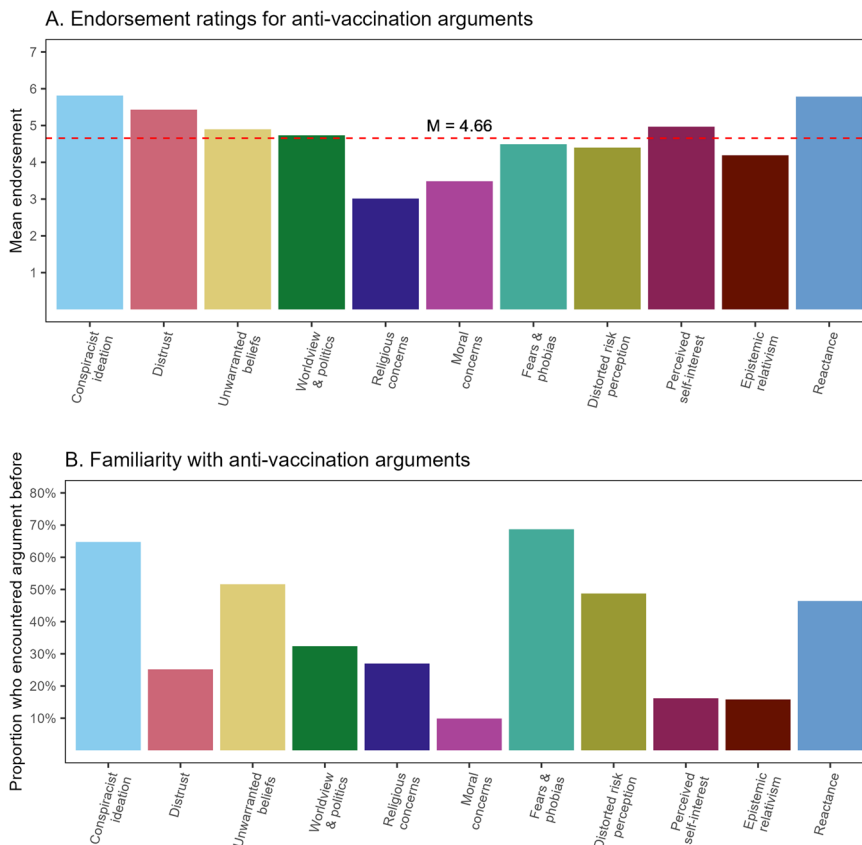
**Participants who endorsed anti-vaccination arguments tended to endorse all.** We first examined participants’ endorsement of the 11 anti-vaccination arguments presented to them, and whether they had encountered those arguments before (see Fig. 2). Participants’ endorsements of all arguments representing different attitude roots were significantly intercorrelated (Cronbach’s  $\alpha = 0.88$ ) and their were not consistently correlated with participants’ familiarity with that argument (see Table S1). We ran 11

multiple linear regressions (one for each argument endorsement) including all the psychological constructs as predictors, which found that the strongest predictor in all models was risk distortion (with the sole exception of centrality of religion for the religious concern argument; see Table 4). Therefore, we replicated previous findings that endorsements appear to indicate the strength of an individual’s beliefs (that lead them to endorse all anti-vaccination arguments, even previously unknown ones), rather than reflecting the source of a person’s hesitancy (Holford et al., 2023).

**Attitude roots in self-expression of anti-vaccination arguments.** While participants might be willing to endorse anti-vaccination arguments that were not necessarily at the root of their hesitancy, we expected that their self-expressed arguments would relate more to people’s attitude roots. We examined participants’ self-expressed texts, which were coded for the expression of all 11 attitude roots (i.e., each text might express one or more attitude roots, each coded as a dichotomous variable; see Fig. 1). Distrust was the most frequently expressed type of argumentation (expressed in 78% of texts), followed by fear and phobias (54%), whereas religious concerns (2%) and perceived self-interest (1%) were most infrequently expressed (see Fig. 3).

**Participants expressed common combinations of arguments when explaining their vaccine hesitancy.** We explored commonalities in the use of argumentation in participants’ self-expression (i.e., in their texts, as illustrated in the first box of Fig. 1). This perspective inspected which combination of arguments participants used in their self-expressed texts to explain their attitudes. We conducted a Latent Class Analysis (LCA) using the polCA package for R (v. 1.6.0.1 Linzer and Lewis, 2011) through JAMOVI (v. 2.3.28). LCA is a person-centred statistical method that identifies subgroups within a larger population, assigning individuals to the most likely class based on their responses to observed categorical variables. We included in this analysis the dichotomous coded variables (expressed/not expressed) for each attitude root in participants’ self-expressed texts, with the exception of the three underrepresented attitude roots (i.e., religious concerns, moral concerns, and perceived self-interest). The reported model here features a three-class solution with the best (lowest) BIC among all models (see Supplementary Material, Table S6 for details on model fit and selection).

Figure 4 illustrates the item response probability (panel A) of the eight attitude roots for each of the three classes identified in the LCA (i.e., the likelihood of expressing that attitude root for that class of participants). We juxtapose this against average scores on the eight matching psychological constructs for each of



**Fig. 2 Average endorsement and familiarity of anti-vaccination arguments.** Panel **A** shows the average endorsement of arguments representing each attitude root and Panel **B** shows the proportion of participants familiar with those arguments.

**Table 4 Number of significant, and strongest, psychological construct predictors in multiple linear regressions on endorsements of arguments that represent each attitude root.**

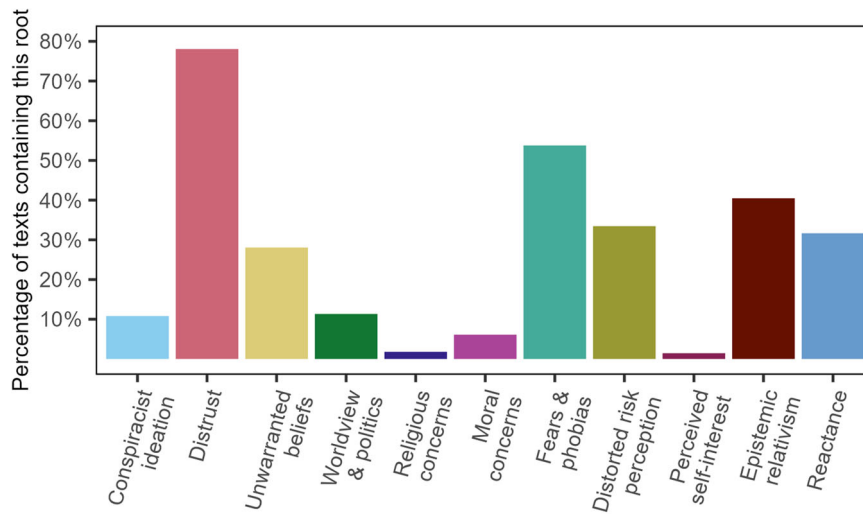
Attitude root	No. of significant predictors	Strongest psychological construct predictor
Conspiracist ideation	4	Risk distortion
Distrust	5	Risk distortion
Unwarranted beliefs	3	Risk distortion
Worldview & politics	4	Risk distortion
Religious concerns	6	Centrality of religion
Moral concerns	5	Risk distortion
Fear & phobias	6	Risk distortion
Distorted risk perception	7	Risk distortion
Perceived self-interest	2	Risk distortion
Epistemic relativism	7	Risk distortion
Reactance	4	Risk distortion

Each row gives the result of one multiple linear regression run for endorsement of an argument representing that attitude root. All regressions included 11 psychological constructs as predictors. Where the strongest predictor is the matching psychological construct for the endorsed argument (i.e., same attitude root), it is given in bold. Full details of this and associated analyses are reported in Tables S2–S5.

these three classes (panel B). The first class included an ‘anti-scientific’ cluster of participants who expressed more arguments related to distorted risk perception, unwarranted beliefs, and epistemic relativism (this class is displayed in red in Fig. 4). The second and third classes both included more expressions of distrust, but had otherwise distinct patterns: the second class

included more expressions of fear-related arguments (yellow), while the third class included politicised expressions of reactance and conspiracist arguments (blue). There is a striking similarity in the patterns in both panels. The anti-scientific class of participants exhibited more risk distortion, politicised participants showed higher levels of reactance, and the fearful participants were higher in general medical anxiety. Some of the differences between panels can be explained by psychometric differences in how the attitude roots are represented in self-expression vs. as a construct. For instance, the psychological construct chosen to represent epistemic relativism (measured by the epistemic belief subscale ‘truth is political’) refers to a politicised, and therefore relativistic, conception of truth. We observed that this variable was more prominent for the politicised profile than for the anti-scientific one.

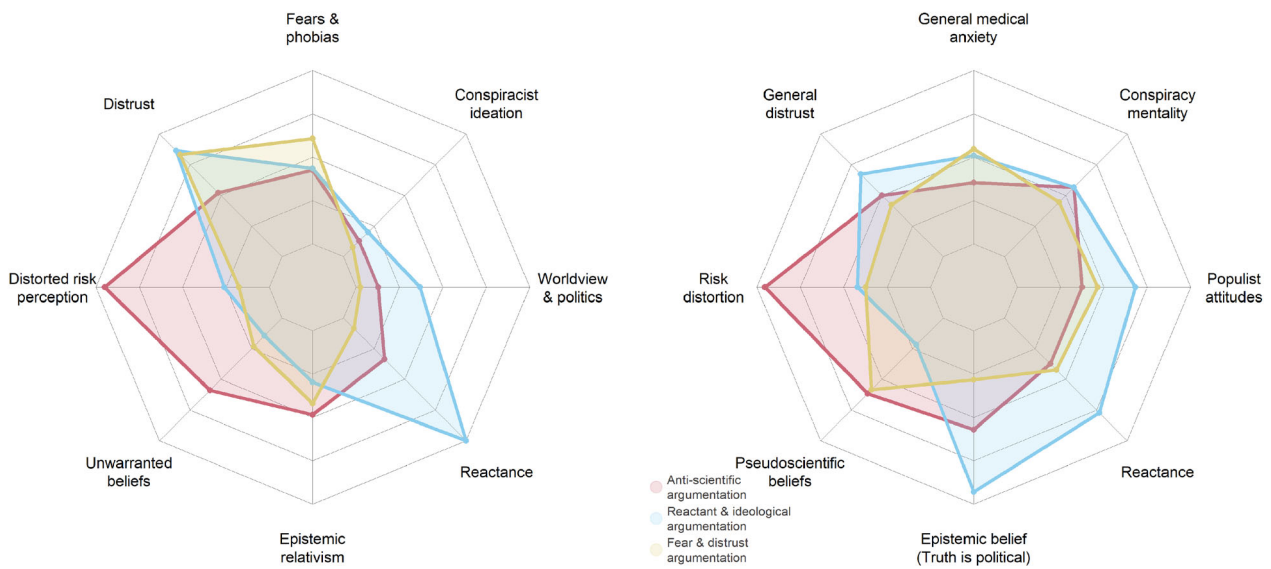
We depict in Fig. 5 the most frequent words that appeared in text from each of the three LCA classes (panel A). Panel B shows the same word frequency illustration at the level of arguments coded as expressing each of the four most prevalent attitude roots. We thus observe from the figure that some of the commonalities in word usage within each LCA class (identified based on attitude roots expressed in whole texts) arise from similar words being used to express the basic units of argumentation for different attitude roots. In other words, some attitude roots are expressed using similar language. This phenomenon was also observed previously by Fasce et al. (2023) and is a common feature in the analysis of natural language patterns, where similarities in linguistic features (vocabulary, semantics, structures, etc.) between different types of expressions can be measured by analysing how often those expressions are confused with each other (Qiu et al., 2020; Rogers et al., 2020). As an example, the language of argumentation for



**Fig. 3 Frequency of self-expression of different attitudes roots in participants' texts.** Each text is coded for the presence of all 11 attitude roots, with one text often containing more than one attitude root. Thus, frequency for each attitude root is calculated as a proportion of all 556 texts and does not sum to 100 across roots.

A. Latent classes in self-expression

B. Mean psychological construct scores for 3 classes

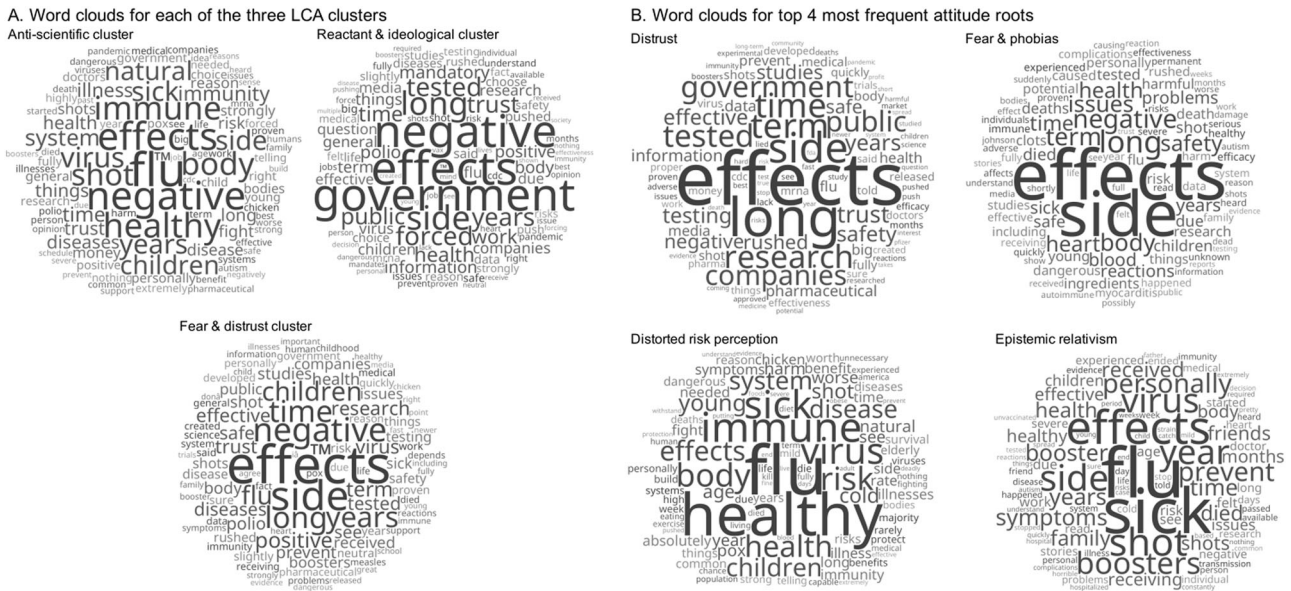


**Fig. 4 Features of the three classes in the LCA.** Panel A shows features of argumentation and panel B shows mean scores on psychological constructs.

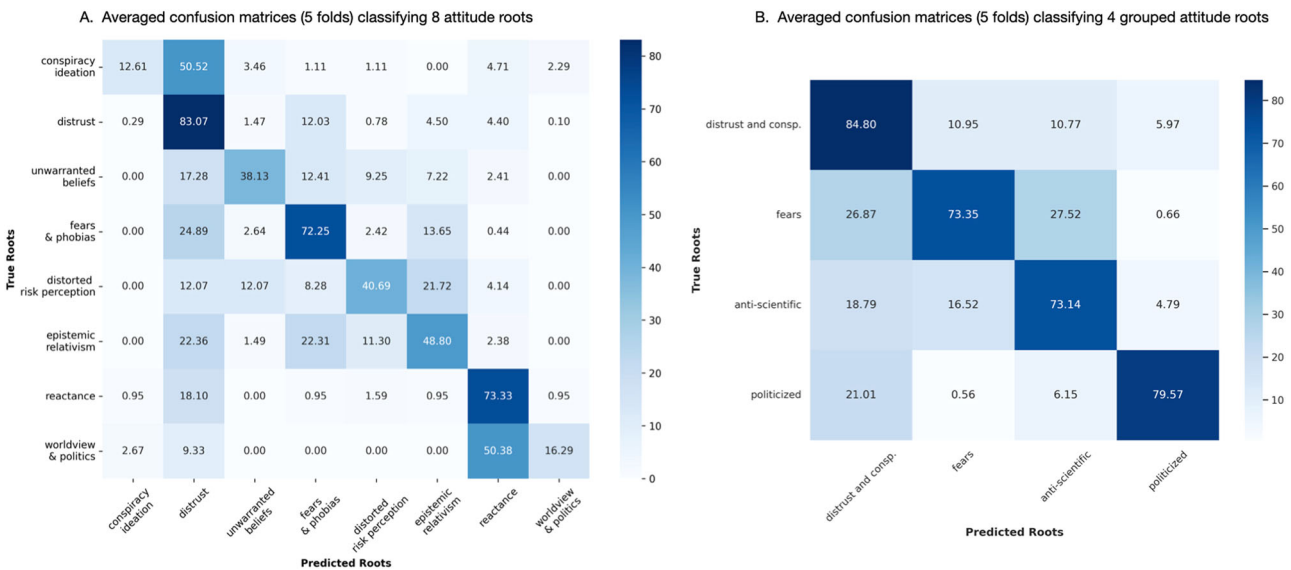
conspiracist ideation shares linguistic features (e.g., vocabulary, semantic space, sentence structure, etc.) with distrust, since people expressing conspiracist arguments may only make very subtle references to suspicions about a plot amidst otherwise distrustful argumentation. One would expect distrust and conspiracist argumentation to be more confused with each other than, say, religious and fear argumentation.

We therefore conducted an exploratory machine learning analysis to assess the extent of such confusability due to linguistic similarities in the self-expression of different attitude roots. In this analysis, we again omitted the three attitude roots (religious concerns, moral concerns, and perceived self-interest) where there was insufficient data for adequate machine training, due to their low frequency among participants' texts. We fine-tuned a multi-label, transformer-based text classifier using DistilBERT<sup>5</sup> to identify the roots of arguments (including multiple roots where

appropriate<sup>6</sup>). In other words, the machine classifier was trained on multiple randomly-selected samples of (labelled) arguments that were segmented from participants' free-text responses and learned to predict whether a new 'unseen' argument (also randomly drawn from the sample of segmented arguments) contained any of the 11 attitude roots. The details of this training and testing are reported in Supplementary Material; our focus was not on maximising predictive performance, but on assessing how the machine classifies different arguments into attitude roots and which roots are confused with one another. Figure 6 plots the machine-predicted roots of arguments in the classification exercise vs. the actual roots for that argument (labelled by coders) as a confusion matrix, showing that the attitude roots distrust, fear and phobias, and reactance, had high percentages of accurate classifications (dark blue values in the diagonal of the matrix). We also observed that conspiracist ideation and



**Fig. 5 Word clouds of arguments.** Panel **A** shows word clouds for each of the three LCA classes. Panel **B** shows word clouds for the four most frequently expressed attitude roots. Word clouds were produced using LIWC-22 software (Pennebaker Conglomerates, 2022). Common English ‘stopwords’ (i.e., words frequently used within the language) and words containing ‘vaccin’, ‘immunis’ and ‘people’ were omitted. We omitted as well “Covid” because of its extremely high incidence across all texts.

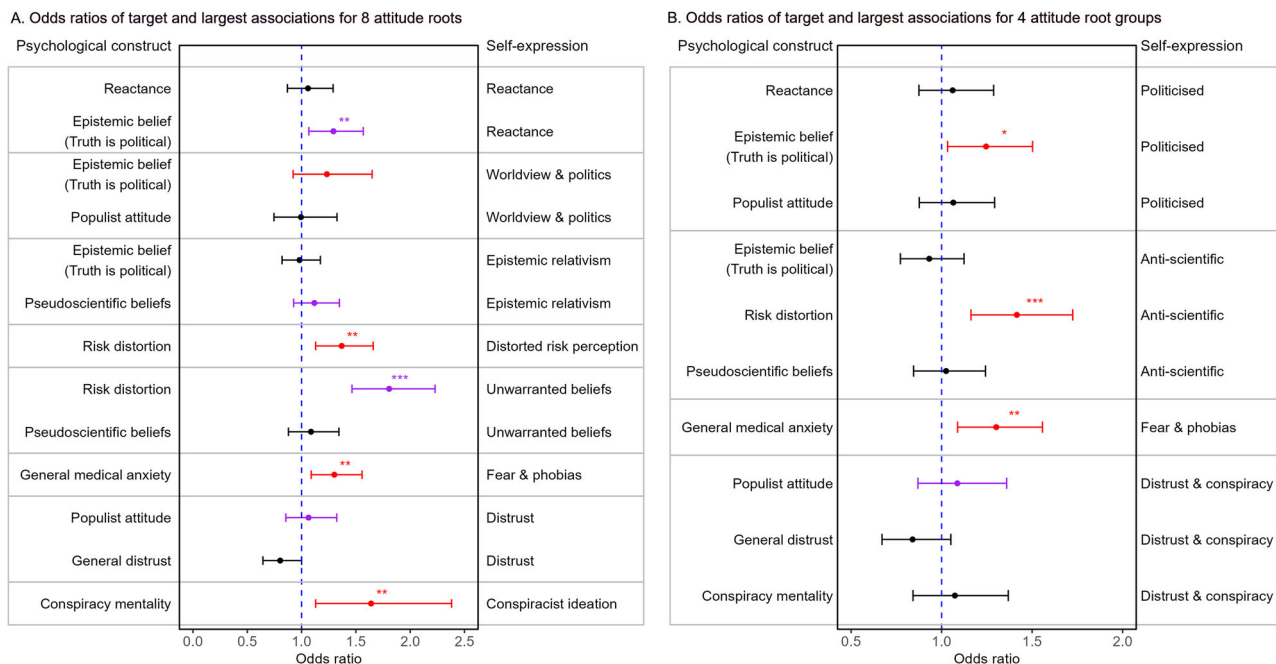


**Fig. 6 Average confusion matrices in classification exercises.** Panel **A** shows confusion matrices for 11 root classification. Panel **B** shows confusion matrices for four grouped roots. The matrices present the average proportion of classified argumentation units across the five cross-validation folds used for training and testing each classification model. Three roots with limited training data to provide meaningful classification results were excluded from the plot. Note that each argumentation unit can have multiple true and predicted classifications to a root/group. F1 scores for the classification are presented in Tables S7, S8.

worldview and politics were frequently classified as distrust and reactance respectively, suggesting that these roots shared substantial linguistic features. The three remaining roots (unwarranted beliefs, distorted risk perception, and epistemic relativism) were consistently confused with each other, but the machine also commonly confused them with distrust and fear. Although this indicates shared linguistic features with distrust and fear, some of that overlap is likely driven by the higher prevalence of these roots in the data (which gives the machine a higher base-rate likelihood of classifying arguments distrust or fear). From a theoretical perspective, it would be reasonable for

unwarranted beliefs, distorted risk perception, and epistemic relativism to share linguistic features related to anti-scientific argumentation.

Based on these observations, we attempted a second classification exercise (Fig. 6, panel B), this time training the machine classifier to predict the category of each argumentation unit as one of four grouped attitude roots on the basis of their shared linguistic features: distrust and conspiracy (combining those two roots), fears and phobias, ‘anti-scientific’ (combining unwarranted beliefs, epistemic relativism, and distorted risk perception), and ‘politicised’ (combining reactance and worldview and



**Fig. 7 Odds ratios obtained from multiple logistic regressions (one per attitude root/grouped roots).** Each grey box presents results from a different multiple logistic regression on a dichotomous self-expressed attitude root (panel A) or grouped roots (panel B) variable, including eight construct predictors in each model (three attitude roots with very low frequency of self-expression are excluded from this model). See Table S9 for results of regressions with all attitude roots. For simplicity of presentation, only the targeted construct predictor(s) and the strongest predictor (if different) are plotted. Red indicates that the targeted construct-self-expression pair is the strongest predictor; purple indicates the strongest predictor in the model that is not one of the targeted pair. Asterisks indicate significant effects in the regression (\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ ).

politics). In this second classification, there was a high proportion of accurate predictions (all >70%), although some linguistic overlap among the groups still remains. Specifically, fear and phobias continued to share linguistic features with the ‘distrust and conspiracy’ and ‘anti-scientific’ groups, while the ‘politicised’ group mainly shared linguistic features with the ‘distrust and conspiracy’ group.

**Self-expressed anti-vaccination arguments are still associated with psychological constructs.** Notwithstanding the substantial overlaps between the language of argumentation across attitude roots, we still expected to find significant positive associations between the self-expression of arguments and psychological constructs that represent the same attitude root. We tested this hypothesis using individual multiple logistic regressions (one per attitude root) on the dichotomous coded variable (expressed or not) for attitude roots in participants’ self-expressed texts. Each regression included all the psychological constructs as predictors. For comparability with the machine analyses in the previous section, and because the low frequency of ‘perceived self-interest’, ‘religious concerns’, and ‘moral concerns’ challenges meaningful interpretation of analyses variables, we report here results for regressions including all eight other roots. Models including all 11 roots are reported in the Supplementary Material, along with pre-registered correlational analyses. The pre-registered analyses produced similar results, but are less robust since they do not control for all psychological constructs. As shown in Fig. 7, panel A (with each grey box representing one regression for a target psychological construct-self-expression pair), three out of eight of the psychological constructs were both significant and the strongest predictor of the expected self-expressed attitude root.

Since we had observed shared linguistic features among four attitude root groupings, we performed the same logistic regression analysis for self-expression belonging to each of the

four groups (i.e., whether participants’ text is coded as a member of that group), again entering all eight psychological construct predictors into the model (Fig. 7, panel B). We found that self-expression belonging to three of four groups was significantly predicted by an associated psychological construct.<sup>7</sup>

Interestingly, we did not find the expected associations for the ‘distrust & conspiracy’ root group. However we note that coders followed an explicit instruction to distinguish between conspiracist ideation and distrust. The strong association between conspiracy mentality and self-expressed conspiracist ideation also reflects that this specific relationship goes beyond one of distrust alone and the linguistic similarities may simply reflect that one will likely express distrust in the course of expressing conspiracist ideation, but not vice versa.

**Discussion**

In this study, we analysed the attitude roots in participants’ self-expression of reasons for negative vaccine attitudes and corresponding measures of psychological constructs, using a framework of 11 attitude roots of anti-vaccination argumentation (Table 1; Fasce et al., 2023). To our knowledge, we are the first to collect primary data on vaccination attitudes that include both these elements. Our study makes three main contributions to attitude root theory, methods for researching it, and implications for identifying attitude roots in practice. First, we illustrate that individuals’ endorsement of anti-vaccination arguments—as measured by the extent they agree with given arguments—are not reliable indicators of their motivations for their negative attitude to vaccination. Rather, to diagnose individuals’ potential attitude root(s), it is necessary to study individuals’ self-expression. Second, we demonstrate that there are commonalities in the ways people express different attitude roots. Looking at the segmented units of argumentation, we observe that arguments coded as expressing different attitude roots nonetheless share linguistic

features. When examining participants' texts, we also saw that certain individuals express common combinations of attitude roots to explain their resistance to vaccines, which mirror their scores on psychological construct measures of those attitude roots (Fig. 4). Third, we identify which attitude roots have a robust and direct association between people's self-expression of the root and their scores on psychological construct measures for that same root (Fig. 7). We discuss our findings in the context of these contributions.

### **Endorsements of given anti-vaccination arguments distinguish one's strength of attitude rather than one's attitude root.**

Participants' endorsements of anti-vaccination arguments presented to them were highly intercorrelated, in line with past findings (Holford et al., 2023). Participants were also more likely to endorse all the anti-vaccination arguments presented to them if they scored higher on most of the psychological measures—which were questionnaires validated in previous research for the constructs they were designed to measure. In addition, the strength of participants' endorsement of anti-vaccination arguments was not consistently related to whether they were familiar with those arguments. Rather, endorsing anti-vaccination arguments seemed to be a reflection of the strength of an individual's negative attitude towards vaccination. We thus replicated the practical limitation of measuring argument endorsements as an indication of individuals' attitude roots: people may be motivated to appraise incoming information on the basis of its alignment with not just their attitude roots, but also the negative attitude itself. Because participants had first written their own explanations of their vaccine attitudes, we already knew what they chose to express. We could thus observe that they endorsed arguments that they had not themselves expressed.

### **Commonalities in self-expression of different attitude roots.**

Another challenge we had identified from previous research (Holford et al., 2023; Hornsey et al., 2018; Soveri et al., 2020), was that individuals could be concurrently high in multiple psychological constructs that motivate anti-vaccination attitudes. For instance, an individual could be motivated by both a tendency to feel others are restricting one's autonomy ('reactance') and a preference for alternative conceptions of facts ('epistemic relativism'). Previous work has shown that many psychological constructs related to vaccine hesitancy do co-occur, with individuals exhibiting high levels of more than one (e.g., Gerace et al., 2022; Goreis and Voracek, 2019; Hornsey et al., 2018; Soveri et al., 2020). A novel aspect of our work was to investigate these overlapping attitude roots within participants' expressed texts. Using LCA, we identified three common ways participants combined different argumentation in their self-expression, featuring (i) anti-scientific, (ii) fearful, and (iii) politicised argumentation. Notably, we also observed that the psychological profiles of these three groups of individuals broadly matched their expressions. In other words, individuals who expressed certain attitude roots in their argumentation tended to also have high levels of the corresponding psychological construct for that attitude root.

We also analysed the language of argumentation itself, using a machine classifier trained on human annotation of whether each unit of argumentation expressed each attitude root. The classifier indicated that the language of expressed fear and phobias was reasonably distinct, but there were strong similarities in the language used to express (i) distrust and conspiracist ideation, (ii) unwarranted beliefs, distorted risk perception, and epistemic relativism (i.e., 'anti-scientific' argumentation), and (iii) world-view and politics and reactance (i.e., 'politicised' argumentation).

Interestingly, this corresponds well to the classes identified in the LCA, despite the two analyses being performed independently and on different levels of text (for the machine classification: single units of argumentation; for the LCA: whole texts expressed by participants). The main difference was that the expression of distrust was more distinctly identified in the machine classification, but spread across LCA classes. This discrepancy between the analyses is unsurprising because distrust was by far the most frequently coded attitude root. Therefore, while it may be consistently identified at the argument level, its appearance in nearly 80% of texts makes it likely to be spread across groups at the text level.

### **Associations of self-expressions with psychological constructs.**

Given the common features among attitude roots when they are expressed in text, as well as the tendency for individuals to score highly on more than one psychological construct of the attitude root, it is remarkable that we were still able to isolate several associations between self-expression and psychological construct. We observed the predicted robust, significant associations between self-expressed attitude root and a matching psychological construct for conspiracist ideation, fear and phobias, and distorted risk perception. These associations thus support our hypothesis, derived from attitude root theory, that one's attitude root (as measured by their levels of a psychological construct) motivates the expression of a particular anti-vaccination argument: someone who has a more general tendency to believe in conspiracies is likely to express a conspiratorial argument, whereas someone who experiences anxiety over medical procedures is likely to express a fear of vaccine side effects.

We then took into consideration the linguistic similarities among expressions of different roots and attempted to predict expression of attitude roots with the same linguistic features (i.e., whether a relevant psychological construct would be associated with expressing one of the linguistically-similar arguments found in the machine classification exercise). Here, we found that risk distortion continued to predict expressions from a group of anti-scientific argumentation (including distorted risk perception, unwarranted beliefs, and epistemic relativism), while expressing beliefs in the politicised nature of truth were also predictive of politicised argumentation expressing the attitude roots of world-view and politics and reactance. Although we had initially selected the scale on the politicised nature of truth as a relativistic measure of epistemic beliefs, one limitation of our study was that due to time constraints, each scale was measured by only three items. In consequence, focusing on the 'truth is political' subset of the epistemic belief scale (Garrett and Weeks, 2017) produced a constructed measure that was more political in nature.

Our approach to predicting expression of linguistically-grouped arguments also had a drawback when it came to distrust & conspiracy-based arguments, which we believe is driven by the high frequency of expressed distrust in our sample. Nearly 80% of texts featured distrust, making it difficult to associate its expression with any one psychological construct. We also note as another limitation constraining our interpretation that our data was from a US sample whose vaccination experiences may be country-specific and not generalisable elsewhere.

### **Practical implications for vaccine communication in the post-Covid period.**

The high frequency of distrust-based argumentation in our data—which was collected about a year after the initial COVID-19 vaccine roll-out in most countries—invites the question of whether post-COVID, attitudes to vaccination have suffered a dip in trust. An interesting observation in our data was that although we asked participants about their attitudes to

vaccination in general, many participants highlighted the COVID-19 vaccine in their responses—to the extent that we had to omit the word ‘COVID’ in our generated word clouds (Fig. 5) so it did not overshadow other words.

Of course, we cannot be certain to what extent participants’ expressed distrust was specific to or motivated by experiences with the COVID-19 vaccine, or if the COVID-19 vaccine simply provided a salient example for argumentation motivated by a general tendency for distrust. However, other work suggests that experiences during the pandemic have had an impact on people’s attitudes. A recent survey of 23 countries found that 23.1% of respondents reported being less willing to receive a vaccination following their pandemic experiences, and 14–17% (depending on the country) reported that they would not get vaccinated if advised to for a new pandemic threat (Lazarus et al., 2024). Trust in science and public health authorities is also increasingly bifurcated along political lines in the US. One study reported an erosion of trust in science over the course of the pandemic among Republicans but not Democrats (Hatton et al., 2022), and there have been concerns that the politicisation of COVID-19 vaccines has influenced ideologically-motivated attitudes toward and uptake of other vaccines (Motta, 2023). Other studies indicate that experiences during the pandemic may have impacted trust in other preventive health information as well, with reduced trust more prominent among minority ethnicity groups compared to majority groups (Bispo et al., 2023). Reports of post-COVID levels of trust in different sources of health information also reveal striking differences across countries in which sources are most trusted (e.g., medical professionals in some cases, religious leaders in others), although sources closest to the individual (e.g., family and friends) tended to rank highest (Lazarus et al. 2024).

These post-COVID patterns of reduced trust in official sources and preferences for sources more congruent to one’s identity necessitate more than ever that public health and medical practitioners communicate in ways that build trust. To do so, practitioners often need to identify individuals’ psychological motivations and demonstrate empathy for these (Gagneur, 2020; Holford et al., 2024; Paterson et al., 2016). We outline here how public health communicators and practitioners could use our findings to help them with their communication challenges.

Our study showed that it is possible to infer what different psychological constructs may motivate individuals’ hesitancy by analysing text that they have produced as spontaneous argumentation against vaccination, but not from agreement ratings of prototypical arguments. This has two important implications for public health teams seeking to monitor vaccine perceptions and attitudes, for example through behavioural insights surveys conducted at local government level (UK Local Government Association, 2021). First, it is preferable to collect and analyse qualitative data, such as open-ended text responses, since quantitative measures, though simpler to analyse, have poor discernment for attitude roots. Improvements in text analysis tools in recent years—including natural language processing techniques and more recently, Large Language Models such as those driving generative artificial intelligence tools—have enabled more nuanced analyses of open-ended text responses to be conducted at scale. Progress in this area suggests that tasks such as the analyses described in this paper could soon be integrated into automatic analysis tools. Future research may wish to investigate how these tools could support healthcare practitioners in providing snapshots of relevant attitude roots at play in people’s anti-vaccination argumentation. Second, when examining factors that are associated with or predict vaccine attitudes, research has often investigated vaccine attitudes as a single construct. Since different types of vaccine attitudes can have different psychological predictors, insights about what is driving

negative vaccine sentiment in an area should pay attention to what type of vaccine attitude is being measured.

Our study also demonstrated how attitude roots can be grouped into meaningful clusters of mutually-reinforcing argumentation, which could be predicted by certain psychological constructs. These findings imply that by asking an individual to explain their attitudes, practitioners can gain insight on that individual’s psychological profile and anticipate the overlapping attitude roots that contribute to the individual’s overall motivations. For example, our results indicate that an individual expressing reactance is likely to be politicised, while an individual who expresses a distorted perception of risk is likely also vulnerable to pseudoscientific beliefs and overreliance on personal testimonies. It is thus possible to anticipate what other arguments the individual may express in conjunction. Knowing this, a practitioner who has identified a few common arguments from their patients can prepare for future interactions by looking up the related attitude roots, associated arguments, and empathetic counter-arguments.<sup>8</sup>

## Conclusion

Our work contributes novel data that goes beyond previous work by linking participants’ self-expressed text explaining their vaccination attitudes to measures of psychological constructs for 11 different attitude roots of anti-vaccination attitudes. Participants often expressed three combinations of attitude roots within their explanations (anti-scientific, fearful, and politicised argumentation), and used similar language to express some attitude roots, but their self-expression broadly reflected their levels of a related psychological construct. Our findings are of practical value to communicators, particularly healthcare professionals, who are encouraged to leverage their position as trusted medical advisors to engage vaccine-hesitant patients in conversation (Dubé, 2017; Moss et al., 2016). Healthcare professionals thus need to understand patients’ motivations from what they express, to be able to tailor their response to address the psychological factors motivating vaccine hesitancy. Our data gives an indication of the frequency of attitude roots individuals may express and several common combinations of arguments that are likely motivated by a specific psychological construct (i.e., attitude root).

## Data availability

All datasets used in this article are publicly available at <https://osf.io/yfks3/>. Source code used in this article is publicly available at <https://osf.io/yfks3/>.

Received: 15 March 2024; Accepted: 27 June 2024;

Published online: 17 July 2024

## Notes

- 1 This political bias in the sample reflects the well-documented political divisions around COVID-19 vaccination in the US at the time of the study (Bolsen and Palm, 2022).
- 2 See Supplementary File 2, Table S10 for wording of arguments. Holford et al. (2023) found a one-factor structure for the agreement with these prototypical arguments, which was also observed in this sample using Confirmatory Factor Analysis [ $\chi^2$  (df) = 213.062 (43),  $p < 0.001$ ; TLI = 0.90; CFI = 0.92; RMSEA (90% CI) = 0.08 (0.073–0.096); SRMS = 0.05; factor loadings > 0.44].
- 3 The decision to segment texts was made to ensure quality and fidelity of the coding process when dealing with texts that contained multiple spans of argumentation. In this case, standard segmentation (e.g., by sentences) was not feasible due to the inconsistency of writing structures among participants.
- 4 The Jaccard index serves as a metric for evaluating sequence similarity (Jaccard, 1901). In the context of our study, this index was applied to sequences of text segments

considered to express argumentation. It thus gauged the similarity in the segmentation patterns of the text, focusing on the overall structure of the segments rather than emphasising the exact starting and ending characters of each segment.

- 5 Language model introduced by Sanh et al. (2020), as a lighter and optimised version of BERT (Devlin et al. 2019).
- 6 In the multi-label process, the model is trained on arguments that may be each annotated with multiple roots, so it predicts all the roots a new argument in the test set may have.
- 7 Although we had intended the epistemic beliefs subscale ('truth is political') to be related to epistemic relativism, we note that it also contains a heavily political dimension, making its association with politicised self-expression logical and theoretically sound.
- 8 A resource that provides information and guidance on crafting such responses and rebuttals to misinformation, based on the same theoretical framework that guided our analyses in this study (Fasce et al., 2023), can be found at <https://jitsuvax.info>.

## References

- Akkerman A, Mudde C, Zaslove A (2013) How populist are the people? Measuring populist attitudes in voters. *Comp Political Stud* 47:1324–1353. <https://doi.org/10.1177/0010414013512600>
- Attwell K, Freeman M (2015) I Immunise: an evaluation of a values-based campaign to change attitudes and beliefs. *Vaccine* 33(46):6235–6240
- Betsch C, Schmid P, Heinemeier D, Korn L, Holtmann C, Böhm R (2018) Beyond confidence: development of a measure assessing the 5c psychological antecedents of vaccination. *PLoS ONE* 13:e0208601
- Bispo JB, Douyon A, Ashad-Bishop K, Balise R, Kobetz E (2023) How trust in cancer information has changed in the era of COVID-19: patterns by race and ethnicity. *J Health Commun* 28:131–143
- Bolsen T, Palm R (2022) Politicization and COVID-19 vaccine resistance in the U.S. *Prog Mol Biol Transl Sci* 188:81
- Boyd RL, Pennebaker JW (2017) Language-based personality: a new approach to personality in a digital world. *Curr Opin Behav Sci* 18:63–68
- Broniatowski DA, Jamison AM, Qi S, AlKulaib L, Chen T, Benton A (2018) Weaponized health communication: Twitter bots and Russian trolls amplify the vaccine debate. *Am J Public Health* 108:1378–1384
- Bruder M, Haffke P, Neave N, Nouripanah N, Imhoff R (2013) Measuring individual differences in generic beliefs in conspiracy theories across cultures: conspiracy mentality questionnaire. *Front Psychol* 4:225
- Chu J, Pink SL, Willer R (2021) Religious identity cues increase vaccination intentions and trust in medical experts among American Christians. *Proc Natl Acad Sci USA* 118(49):e2106481118
- Council of Canadian Academies Fault Lines (2023) Expert Panel on the Socio-economic Impacts of Science and Health Misinformation
- Devlin J, Chang M-W, Lee K, Toutanova K (2019) BERT: pre-training of deep bidirectional transformers for language understanding. In Burstein J, Doran C, and Solorio T (eds) *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pp 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics
- Dubé E (2017) Addressing vaccine hesitancy: the crucial role of healthcare providers. *Clin Microbiol Infect* 23:279–280
- Ecker UKH, Lewandowsky S, Cook J, Schmid P, Fazio LK, Brashier N (2022) The psychological drivers of misinformation belief and its resistance to correction. *Nat Rev Psychol* 1:13–29
- Fasce A, Avendaño D, Adrián-Ventura J (2021) Revised and short versions of the pseudoscientific belief scale. *Appl Cogn Psychol* 35:828–832. <https://doi.org/10.1002/acp.3811>
- Fasce A, Schmid P, Holford DL, Bates L, Gurevych I, Lewandowsky S (2023) A taxonomy of anti-vaccination arguments from a systematic literature review and text modelling. *Nat Hum Behav* 7(9):1462–1480
- Gagneur A (2020) Motivational interviewing: a powerful tool to address vaccine hesitancy. *Can Commun Dis Rep*. 46:93–97
- Garrett RK, Weeks BE (2017) Epistemic beliefs' role in promoting misperceptions and conspiracist ideation. *PLoS ONE* 12:e0184733–
- Gerace A, Rigney G, Anderson JR (2022) Predicting attitudes towards easing COVID-19 restrictions in the United States of America: the role of health concerns, demographic, political, and individual difference factors. *PLoS ONE* 17(2):e0263128
- Gerbasi ME, Prentice DA (2013) The self- and other-interest inventory. *J Personal Soc Psychol* 105(3):495–514
- Goreis A, Voracek M (2019) A systematic review and meta-analysis of psychological research on conspiracy beliefs: field characteristics, measurement instruments, and associations with personality traits. *Front Psychol* 10:425400
- Greenwood B (2014) The contribution of vaccination to global health: past, present and future. *Philos Trans R Soc B: Biol Sci* 369:20130433
- Hatton CR, Barry CL, Levine AS, McGinty EE, Han H (2022) American trust in science & institutions in the time of COVID-19. *Daedalus* 151(4):83–97
- Holford DL, Fasce A, Costello TH, Lewandowsky S (2023) Psychological profiles of anti-vaccination argument endorsement. *Sci Rep* 13(1):11219
- Holford DL, Schmid P, Fasce A, Lewandowsky S (2024) The Empathetic Refutational Interview to tackle vaccine hesitancy: four randomised experiments. *Health Psychol* 43:426–437
- Hong S-M, Faedda S (1996) Refinement of the Hong psychological reactance scale. *Educ Psychol Meas* 56:173–182. <https://doi.org/10.1177/0013164496056001014>
- Hornsey MJ (2020) Why facts are not enough: understanding and managing the motivated rejection of science. *Curr Dir Psychol Sci* 29:583–591. <https://doi.org/10.1177/0963721420969364>
- Hornsey MJ, Harris EA, Fielding KS (2018) The psychological roots of anti-vaccination attitudes: a 24-nation investigation. *Health Psychol* 37:307–315
- Huber S, Huber OW (2012) The centrality of religiosity scale (crs). *Religions* 3:710–724
- Jaccard P (1901) Étude comparative de la distribution florale dans une portion des Alpes et du Jura. *Bull Société Vaud Sci Nat* 37:547–579
- Karlsson LC, Mäki KO, Holford D, Fasce A, Schmid P, Lewandowsky S, et al. (2024) Testing psychological inoculation to reduce reactance to vaccine-related communication. *Health Commun* 7:1–9
- Lasser J, Aroyehun ST, Carrella F, Simchon A, Garcia D, and Lewandowsky S (2023) From alternative conceptions of honesty to alternative facts in communications by US politicians. *Nat Hum Behav* 7:1–12
- Lazarus JV, White TM, Wyka K, Ratzan SC, Rabin K, Larson HJ, et al. (2024) Influence of COVID-19 on trust in routine immunization, health information sources and pandemic preparedness in 23 countries in 2023. *Nat Med* 30:1559–1563
- Lewandowsky S, Oberauer K (2016) Motivated rejection of science. *Curr Dir Psychol Sci* 25:217–222
- Linzer DA, Lewis JB (2011) polca: an R package for polytomous variable latent class analysis. *J Stat Softw* 42:1–29
- Moss JL, Reiter PL, Rimer BK, Brewer NT (2016) Collaborative patient-provider communication and uptake of adolescent vaccines. *Soc Sci Med* 159:100–107
- Motta M (2023) Is partisan conflict over COVID-19 vaccination eroding support for childhood vaccine mandates? *npj Vaccines* 8(1):1–4
- Noecker Jr J, Ryan M, Juola P (2013) Psychological profiling through textual analysis. *Lit Ling Comput* 28(3):382–387
- Paterson P, Meurice F, Stanberry LR, Glismann S, Rosenthal SL, Larson HJ (2016) Vaccine hesitancy and healthcare providers. *Vaccine* 34:6700–6706
- Pennebaker Conglomerates (2022)
- Peterson B, Smith JA, Tannenbaum D, Shaw MP (2009) On the “exporting” of morality: its relation to political conservatism and epistemic motivation. *Soc Justice Res* 22:206–230
- Qiu X, Sun T, Xu Y, Shao Y, Dai N, Huang X (2020) Pre-trained models for natural language processing: a survey. *Sci China Technol Sci* 63(10):1872–1897
- Rogers A, Kovaleva O, Rumshisky A (2020) A primer in BERTology: what we know about how BERT works. *Trans Assoc Comput Linguist* 8:842–866
- Ruggeri K, Vanderslott S, Yamada Y, Argyris YA, Veckalov B, Boggio PS (2024) Behavioural interventions to reduce vaccine hesitancy driven by misinformation on social media. *BMJ* 384:e076542
- Sanh V, Debut L, Chaumond J, and Wolf T (2020) DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter
- Schmid P, Rauber D, Betsch C, Lidolt G, Denker M-L (2017) Barriers of influenza vaccination intention and behavior—a systematic review of influenza vaccine hesitancy, 2005 - 2016. *PLoS ONE* 12:e0170550
- Simchon A, Sutton A, Edwards M, Lewandowsky S (2023) Online reading habits can reveal personality traits: towards detecting psychological microtargeting. *PNAS Nexus* 2:pgad191
- Soveri A, Karlsson LC, Mäki O, Antfolk J, Waris O, Karlsson H (2020) Trait reactance and trust in doctors as predictors of vaccination behavior, vaccine attitudes, and use of complementary and alternative medicine in parents of young children. *PLoS ONE* 15:e0236527
- Soveri A, Karlsson LC, Mäki KO, Holford D, Fasce A, Schmid P, et al. (2023) Trait reactance as psychological motivation to reject vaccination: two longitudinal studies and one experimental study. *Appl Psychol: Health Well-Being* 16:597–614
- UK Local Government Association (2021) Applying behavioural insights to improve COVID vaccination uptake: a guide for councils
- van Balen F, Verdurmen J (1999) Medical anxiety and the choice for treatment: the development of an instrument to measure fear of treatment. *Psychol Health* 14(5):927–935
- Verma V, Aggarwal RK (2020) A comparative analysis of similarity measures akin to the Jaccard index in collaborative recommendations: Empirical and theoretical perspective. *Soc Netw Anal Min* 10(1):43
- Yamagishi T (1988) The provision of a sanctioning system in the United States and Japan. *Soc Psychol Q* 51:265–271

## Author contributions

DH: Conceptualisation, Data curation, Formal analysis, Investigation, Methodology, Project administration, Validation, Visualisation, and Writing—original draft. EL:

Conceptualisation, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualisation, Writing—original draft, and Writing—review & editing. AF: Conceptualisation, Data curation, Formal analysis, Investigation, Methodology, Writing—original draft, and Writing—review & editing. LCK: Data curation, Investigation, Methodology, Visualisation, and Writing—review & editing. SL: Conceptualisation, Funding acquisition, Methodology, Supervision, and Writing—review & editing.

### Competing interests

All authors declare no competing interests. DH, AF, LK, and SL received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 964728 (JITSUVAX). EL and SL received funding from Deutsche Forschungsgemeinschaft (DFG) - Project number 458366841 (POLTOOLS - Assisting behavioural science and evidence-based policymaking using online machine tools). SL was also supported by an award from the Humboldt Foundation.

### Ethical approval

Ethical approval for the study was obtained from the University of Bristol's Research Ethics Committee (reference: 10708) prior to data collection. The study was carried out in accordance with the approved procedures and in line with the principles of the Declaration of Helsinki.

### Informed consent

Informed consent was obtained from all participants at the point where they were invited to complete the study, using the online survey software that administered the study. Participants gave consent after reading the study information by completing an online tick-box that ensured their anonymity.

### Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1057/s41599-024-03416-4>.

**Correspondence** and requests for materials should be addressed to Dawn Holford.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024