The Reasoning through Evidence versus Authority (EvA) Scale: Scale Development and Validation

Hwayong Shin[1], Priti Shah[2], and Stephanie D. Preston[2]

[1] Department of Political Science, University of Michigan

[2] Department of Psychology, University of Michigan

October 27, 2022

Author Note

This research was supported by a generous grant from the Martin and Ruth Jaffe Fund to Stephanie D. Preston. We thank research assistance from Tyler Hill and Nicholas Hollman with the initial scale development and pilot study, and Hannah Groenke and Amanda Szczesniak with conducting a survey and organizing results for Study 2.

This work involves no conflicts of interests and was approved by the university Institutional Review Board. Our data, code, and study materials are available at: https://osf.io/qeav5/?view_only=cc20ea01c47b48adbde31f7e1e7ec52d. The preregistration for Study 3 is available at: https://aspredicted.org/RTS_PWL. The preregistration for Study 4 is available at: https://aspredicted.org/9HC_B7G.

Correspondence concerning this article should be addressed to Hwayong Shin, Department of Political Science, University of Michigan, 505 S State Street, Ann Arbor, MI 48109-1045, Email: hwayong@umich.edu.

Abstract

Our well-being can improve when people heed evidence rather than simply follow familiar or charismatic authorities who neglect evidence. We developed the Reasoning through Evidence versus Authority (EvA) scale to measure individual differences in reasoning through evidence like science and statistics versus following authority figures such as politicians and celebrities. No existing scales directly measure these tendencies; moreover, it was theoretically unknown whether they reflect a single dimension (from evidence- to authority-based) or separate tendencies to value or distrust each. Our scale validation process included qualitative interviews, and four studies that involved 1583 respondents (753 college graduates, 830 non-college graduates) in which we conducted exploratory and confirmatory factor analyses, and tests of convergent validity, discriminant validity, and measurement invariance by gender and education. This process yielded a 16-item EvA scale with four dimensions: Pro-evidence, Anti-evidence, Pro-authority, and Anti-authority. In assessing criterion validity, these tendencies identified individual differences in important, real-world attitudes and behaviors, including susceptibility to health misinformation, adherence to CDC guidelines on social distancing, confidence in the COVID vaccine, science curiosity, and religiosity. The EvA scale extends our understanding of individual differences in reasoning tendencies that shape critical attitudes, decisions, and behaviors and can help promote informed decisions.

Keywords: reasoning tendency, evidence, authority, scale development

The Reasoning through Evidence versus Authority (EvA) Scale: Scale Development and Validation

Before the introduction of the measles vaccine in 1963, an estimated 30 million people worldwide were infected with measles and more than 2 million died from measles each year; an estimated 50,000 hospitalizations occurred annually in the United States alone (Rota et al., 2016). Subsequently, mass immunization through the measles vaccine dramatically reduced the number of cases, preventing an estimated 17.1 million deaths between 2000 and 2014 worldwide (Perry et al., 2015). Despite the well-documented evidence in support of the measles vaccine, some still refuse the vaccine for non-scientific reasons, such as a general distrust in science or because they follow anecdotal claims (e.g., celebrities' claims that vaccines cause autism) or religious doctrines (e.g., porcine or bovine components of the measles vaccine are prohibited in some religions) (Browne et al., 2015; Martinez-Berman et al., 2020; Wombwell et al., 2015). The resulting avoidance of vaccination has led to measles outbreaks amongst several intentionally unvaccinated communities, threatening the health of vaccinated individuals and those who cannot receive the vaccine, and imposing a burden upon society (Nelson, 2019; Tanne, 2019).

Public health crises, and society overall, benefit when individuals actively seek and base decisions upon "evidence" over "authority"—two major ways of knowing. Evidence refers to aggregated data that describes observed relationships, acquired through the scientific method and statistics (Brown et al., 2010; McNeill & Martin, 2011). Evidence-based reasoning involves drawing conclusions from the best available evidence, obtained through transparent and reproducible procedures, which reduces misleading biases (Gambrill, 1999; Jamieson et al., 2019). In contrast, authority-based reasoning is characterized by arriving at conclusions through

others' opinions, declarations of 'authorities,' unchecked institutions, anecdotes, and popularity, which may lead to the uncritical acceptance of authority opinions that counter evidence (Gambrill, 1999; Guzelian & Guzelian, 2004). Thus, evidence-based practices and decisions have been recommended over authority-based ones in multiple fields, including medicine, management, and policy, to improve outcomes (Akobeng, 2005; Sanderson, 2002; Pfeffer & Sutton, 2006). While evidence is the preferred basis for knowing and deciding, some still believe scientifically unsupported claims, even when made aware of established scientific evidence (Kahan, 2015; Rynes et al., 2018). Instead, they often base decisions upon non-expert authorities, including politicians, celebrities, religious leaders, and friends (Druckman et al., 2013; Martinez-Berman et al., 2020; Waldinger, 2004).

Understanding people's reliance upon evidence and authority is important in the current polarized and fragmented information environment, where individuals can selectively accept or avoid either source of information. Partisan polarization in the US, where Democrats and Republicans are increasingly at odds with and hostile to one another (Abramowitz & Webster, 2016; Iyengar et al., 2019), can motivate individuals to heed their preferred authority figures over evidence. For instance, people are vulnerable to falsehoods or conspiracy theories from well-known politicians and likely to be skeptical of scientific findings resisted by their party's authorities (Bolsen & Druckman, 2018; Miller et al., 2016). Furthermore, people are exposed to both credible information from experts and falsehoods from unreliable sources, making it more important than ever to seek and employ objective evidence to inform decisions and prevent false beliefs (Bronstein et al., 2019; Vosoughi et al., 2018).

It is imperative that we understand people's general tendencies to value or devalue evidence and authorities, which can prevent or facilitate misperceptions and ill-informed

decisions. As a first step toward the systematic study of these reasoning tendencies, we developed the Reasoning through Evidence versus Authority (EvA) scale, to measure individual differences in the tendency to seek versus suspect evidence and to reply upon or resist authority opinions when making decisions.

Measuring Reasoning through Evidence and Authority

There are no existing scales that measure tendencies toward reasoning through evidence and authority. There are related scales such as Need for Cognition (Cacioppo et al., 1984), which measures the degree that people enjoy deliberation and effortful thinking. There are also scales that measure people's belief and trust in science and scientists (Farias et al., 2013; Nadelson et al., 2014). Other related scales measure people's belief in hierarchical order, such as Right-Wing Authoritarianism (Altemeyer, 1998) and general Dispositional Trust (Bianchi & Brockner, 2012). While these constructs overlap with our main interest, none directly address how people seek, employ, or avoid evidence versus authorities as sources of information when making decisions. For example, Need for Cognition captures one's willingness to exert cognitive reasoning efforts, but not how they approach or avoid sources of information (e.g., scientific research, authorities). Belief in Science measures support for scientific evidence per se, without measuring a potentially distinct tendency to discredit science. None of these scales measure one's tendency to follow the advice of authorities when forming beliefs or making decisions. Right-Wing Authoritarianism measures support for authorities and a hierarchical social order—including the government, the elderly, laws, and God, without considering other prominent advisors like friends, celebrities, bloggers, and politicians that may considerably influence people. Dispositional Trust addresses trust or suspicion of others generally, without specifically addressing science or authorities. Neither Right-Wing Authoritarianism nor Dispositional Trust

are directly contrasted with trust in evidence. Thus, existing scales address aspects of the problem we seek to solve, but none directly measure people's tendency to rely upon evidence versus authorities for decisions.

This research is also theoretically important because it is unknown whether support for evidence and authorities rely upon common or distinct constructs. There could be one underlying tendency to base decisions upon evidence on one end of the spectrum to authorities on the other. Alternatively, people's reliance upon evidence or authorities could depend upon separate dimensions, each of which could include seeking versus distrusting those sources. For example, people can separately vary on how much they *seek* and *suspect* evidence. Individuals who value evidence may also accept the advice of authorities. Because people's reasoning may not involve a simple trade-off between evidence versus authority, we considered four potentially separable dimensions: tendencies to seek versus distrust evidence and to rely upon versus suspect authorities. To examine the underlying conceptual structure, we included items representing each of these four possible constructs in developing the scale.

We developed a scale that measures individuals' tendencies to reason through Evidence versus Authority (EvA) in a multi-step process: 1) developing an initial set of items refined through qualitative, in-person interviews; 2) performing exploratory factor analysis to identify underlying dimensions; 3) performing confirmatory factor analysis and establishing measurement invariance across demographic subgroups; 4) assessing convergent and discriminant validity by comparing EvA tendencies to related existing scales; 5) assessing criterion validity by examining behaviors or attitudes related to EvA tendencies.

Study 1: Exploratory Factor Analysis and Initial Scale Characterization

Study 1 was designed to explore the underlying dimensions of reasoning tendencies through evidence and authorities from our initial 57 items. In this study and the following ones, we recruited a similar number of participants with and without a college degree, because of the widespread assumption that education level is positively correlated with scientific reasoning but negatively correlated with non-scientific reasoning, such as conspiratorial thinking (Huber & Kuncel, 2015; van Prooijen, 2017).

Pilot Study

We first generated a pool of 57 items to measure individuals' tendencies to reason from evidence (e.g., scientific research, statistics) or authorities (e.g., parents, friends, politicians, celebrities, religious leaders), with a similar number of items for each dimension (17 Pro-evidence, 12 Anti-evidence, 13 Pro-authority, and 12 Anti-authority; Table S3). Some items were adapted from relevant scales that addressed our concepts of interest: Schommer Epistemological Questionnaire (5 items; Schommer, 1998), Epistemic Beliefs Inventory (1 item; Schraw et al., 2002), Updated Dogmatism Scale (4 items, Shearman & Levine, 2006), and Attitudes Toward Science Scale (2 items; ATSS, Francis & Greer, 1999). The remaining items were generated by our study team to create a range of statements assessing people's reasoning through evidence versus authorities in a variety of relevant domains. From in-person interviews (5 undergraduate students, 4 adults without college education), four items were clarified (Table S1) but none were removed. Details about the procedures and findings from the pilot study are available in supplementary materials.

Methods and Materials

Following the guidelines on the minimum ratios of participants to items (5:1 or 10:1) (Gorsuch, 1983; Worthington & Wittaker, 2006), we recruited 579 participants via

CloudResearch—a survey platform that recruits subjects from Amazon Mechanical Turk (Litman, 2017). For data quality, we followed the Mturk acceptance ratio of 95%, which produces comparable results to laboratory-based studies (Johnson & Borden, 2012). We excluded 28 who failed at least one of four attention checks, which can identify low-quality responses (Berinsky et al., 2019). Two respondents who did not complete the scale were excluded (Newton et al., 2021). 549 respondents (college: 243; non-college: 306) were retained for analysis (demographics in Table S2), leaving an acceptable participant-item ratio of 9.6:1. Participants responded to the 57 items for the EvA scale (order randomized) before completing demographic items (e.g, gender, education).

Results

Exploratory Factor Analysis

We conducted exploratory factor analysis (EFA) with maximum likelihood (ML) estimation to examine the underlying structure of items (Barker et al., 2010; Fabrigar et al., 1999). Our Kaiser-Meyer-Olkin (KMO) was .91, which exceeded the recommended values of .60 and higher for an adequate sample size for EFA (Beavers et al. 2013; Tabachnick & Fidell, 2001). Because EvA reasoning tendencies are best characterized as correlated with each other rather than orthogonal, we used oblique (promax) rotation (Brown, 2015; Fabrigar et al., 1999).

The scree plot and parallel analysis suggested six factors (Figure S1; Cattell, 1966; Hayton et al., 2004). The results of EFA indicated six factors explaining 27%, 24%, 15 %, 14 %, 12%, and 9% of the variance, respectively. Following recommended item deletion criteria (Baker et al., 2010; Haws et al., 2012; Worthington & Whittaker, 2006), eight items were dropped due to cross-loading (> .30), ten items were dropped due to weak factor loading (< .40), and none were dropped due to low communalities (< .40) (Details including item deletion criteria, full

EFA results, with factor loadings, kurtosis, skewness, and communalities are in supplementary materials, Table S6.)

The last step of EFA was to shorten the scale and retain a similar number of items per factor (Baker et al., 2010) to increase scale efficiency (e.g., respondent fatigue) while retaining internal consistency (Worthington & Whittaker, 2006). Empirical and substantive rationales were applied to select a sensible set (Worthington & Whittaker, 2006). To balance factors, we reduced each dimension to four items, deleting those with the following properties: low loadings, high cross-loadings, low contribution to internal consistency, and low conceptual consistency with other items (Brown, 2015; Worthington & Whittaker, 2006) (Table S7). The retained four items per six dimensions are provided in Table S8.

All six factors could potentially be retained, but conceptual interpretability and theoretical relevance are also important for factor retention decisions (Worthington & Whittaker, 2006). We reasoned that it was sensible to drop Factor 6 (Anti-evidence 2), because it was specific to the aversion to medicine and chemicals. Among the two factors related to Pro-authority, we removed Factor 3 and kept Factor 5, because some Factor 3 items measured Anti-authority tendencies and most originated from existing scales, whereas Factor 5 comprised our novel items, contributing more as an original scale. Factor 5 items were also more relevant to our goal of assessing individuals' tendency to rely upon authority figures that they like and follow, including politicians, celebrities, and friends. Conversely, Factor 3 items comprised only hierarchical or traditional authorities (e.g., law, government, God) that are less relevant to our broader concern with recent trends in information sources. The final 16-item scale (item wordings in Table 1) was efficient and balanced, with four factors, each containing four items.

Confirmatory Factor Analysis

Following the convention to employ model fit statistics to compare alternative models after EFA (Baker et al. 2010; Cassidy et al., 2005; Svedholm-Häkkinen & Lindeman, 2017), we assessed the soundness of our factor selection decision compared to alternative models through confirmatory factor analysis (CFA). For CFA, the variance–covariance matrices were analyzed using latent variable software programs and maximum-likelihood minimization functions (lavaan in R, Rosseel, 2012). We evaluated model fit following recommended criteria: RMSEA and SRMR ≤ .08, CFI and TLI ≥ .90 (Bentler, 1990; Byrne, 1994; Fabrigar et al. 1999; McDonald & Ho, 2002). The proposed four-dimensional model had a good fit, $\chi^2(df) = 228.89$ (98), RMSEA = .049, SRMR = .046, CFI = .954, TLI = .943 (Table 1). Item-total correlations indicated that all items contributed to scale homogeneity (.40-.74; Streiner et al., 2015; e.g., Duckworth et al., 2007; Lipkus et al., 2001).

Table 1

Confirmatory Factor Analysis and Item-total Correlations

	Factor loadings	Item-total correlation
Pro-evidence		
When I hear a news story reporting research about health, I want to look up the study they are referring to.	.70	.63
I carefully examine research on important issues to make sure it is valid and unbiased.	.76	.68
When debating an important issue, I try to fact-check things that people state as statistics	.72	.61
When someone makes a statement that sounds like a fact, I want to know the evidence behind it.	.78	.67
Anti-evidence		
Scientific evidence is overrated; there are often better ways to understand the world.	.77	.70
Even if scientific studies are done carefully and transparently, I still don't really believe them.	.81	.73
People make too much of scientific studies in the news when I know that the research is biased anyway.	.78	.68
When new evidence reverses a previous scientific theory, I just stop paying attention to it and make my own decisions.	.72	.64
Pro-authority		
When I have to vote, I see what my politician says and follow their lead.	.65	.53
I often make changes to my diet based on what my friends tell me is more healthy.	.53	.42

When I think a politician has a confident, assertive personality, I naturally like them and vote for them.	.60	.45
I assume that when my favorite blogger or social media personality gives advice, they know what they are talking about.	.59	.44
Anti-authority		
I am concerned that news reports are based on people's opinions rather than actual evidence.	.72	.59
Government officials often say things that are untrue in their public statements.	.66	.57
Hosts of major television news shows do not know enough to be reliable sources of information.	.69	.55
People who are telling us how to act don't always have an incentive to tell the truth.	.50	.42
CFA fit statistics		
CFI	.954	
TLI	.943	
SRMR	.046	
RMSEA	.049	
$\chi^2(df)$	228.89 (98)	
N	547	

Note. Entries for factor loadings are standardized and all were statistically significant ($p < .01$).

We examined the descriptive goodness-of-model fits of our proposed 4-factor solution (Factors 1, 2, 4, 5), compared to the 6-factor solution (all factors retained), a 5-factor solution (Factors 6 also retained), and a 4-factor solution (Factor 3, rather than 5, represents Pro-Authority). Compared to the proposed model, alternative models had worse fits, indicated by greater RMSEA and SRMR or smaller CFI and TLI (Table 2). The proposed and alternative models were non-nested (i.e., did not share the same set of parameters), so we additionally compared models through the Akaike Information Criterion (AIC; after Schermelleh-Engel et al., 2003). The proposed model had a lower AIC, suggesting a better model fit compared to alternatives. These results corroborated our item and factor selections.

Table 2

Fit Statistics for Alternative Models (Number of Factors) for the EvA scale

	$\chi^2(df)$	RMSEA	SRMR	CFI	TLI	AIC
Proposed 4-dim model	228.89 (98)	.049	.046	.954	.943	27851.28
Alternative 6-dim model	687.43 (237)	.059	.068	.901	.884	43233.08
Alternative 5-dim model	404.01 (160)	.053	.050	.932	.920	35693.44
Alternative 4-dim model	351.33 (98)	.069	.073	.920	.902	27937.55

Note. Proposed 4-dimensional model: Four-factor solution with Factors 1, 2, 4, 5; *Alternative 6-dim:* Six-factor solution with six factors (Factors 1, 2, 3, 4, 5, 6); *Alternative 5-dim:* Five-factor solution with five factors (Factors 1, 2, 4, 5, 6); *Alternative 4-dim:* Four-factor solution with Factors 1, 2, 3, 4); dim = dimension.

Factor correlations supported our decision to use oblique rotation, which treats factors as distinct but correlated (–.22-.36, Table S9). Correlations among EvA subscales suggested neither poor discriminant validity nor a single, higher-order factor. None were above .80, thus not suggesting poor discriminant validity, and these correlations varied across factors, suggesting that higher-order factors were unlikely (Brown, 2015). The four EvA constructs had acceptable internal reliability with Cronbach's alpha (.68-.85, Table S9; Bland & Altman, 1997; DeVellis, 2017; Tavakol & Dennick, 2011).

Study 2: Confirmatory Factor Analysis on the EvA Scale

The purpose of Study 2 was to assess the factor structure of the EvA scale by conducting confirmatory factor analysis (CFA) on the 16 items from Study 1 on an independent sample.

Methods

We recruited 201 US adults through CloudResearch. We excluded 12 participants who missed at least one of three attention checks, leaving 189 for analysis (college: 88; non-college: 101; demographics in Table S3).

Results

The proposed 16-item, four-factor EvA scale was supported by CFA, with acceptable model fit and strong factor loadings ($\chi^2(df)$ = 172.349 (98), RMSEA = .063, SRMR = .069, CFI = .928, TLI = .911; Table S10). We explored the structure of the EvA scale by examining four plausible alternative models (after Cable & DeRue, 2002; Medsker et al., 1994). We tested: A) a

more restricted version of our four-factor model, wherein the EvA factors are unrelated to one another (orthogonal), keeping all other specifications the same; B) a model that assumes only one EvA tendency across all four factors; C) a model that assumes two dimensions: Evidence-oriented reasoning (Pro-evidence, Anti-authority) and Authority-oriented reasoning (Pro-authority, Anti-evidence) with a two-factor solution with eight items each; D) a hierarchical model with two second-order factors, one related to evidence-oriented (linked to Pro-evidence and Anti-authority first-order factors) and another to authority-oriented (linked to Pro-authority and Anti-evidence first-order factors) tendencies. Chi-square difference tests (Brown 2015; Schermelleh-Engel et al., 2003) indicated a significantly better model fit for our proposed model compared to alternatives (Table 3; *p*s < .01). All four alternatives did not meet the criteria for an acceptable fit with respect to all model fit indices, indicating that it is more likely that there are four factors for accepting and rejecting both evidence and authority, which are related and unlikely to reflect a single dimension, two combined factors, or two second-order factors. These results suggest convergent and divergent validity of four related yet distinct EvA subscales: Anti-evidence, Pro-evidence, Anti-authority, and Pro-authority.

Table 3

Fit Statistics for the Proposed and Alternative Models

	$\chi^2(df)$	$\chi^2_{diff}(\Delta df)$	RMSEA	SRMR	CFI	TLI
Proposed model	172.35 (98)		.063	.069	.928	.911
Alternative model A	293.67 (104)	121.32 (6)***	.098	.149	.815	.787
Alternative model B	763.39 (104)	591.04 (6)***	.183	.194	.358	.259
Alternative model C	428.89 (103)	256.54 (5)***	.129	.134	.683	.630
Alternative model D	243.85 (103)	71.5 (5)***	.085	.106	.863	.840

Note. *Proposed model:* Items load on four factors (Pro-evidence, Anti-evidence, Pro-authority, Anti-authority); *A:* No relationships between factors; *B:* All items load on one factor; *C:* Items load on two factors (Evidence-oriented, Authority-oriented); *D:* Items load on four first-order

factors, with two second-order factors (Evidence-oriented, Authority-oriented). **$p < .1$; **$p < .05$; ***$p < .01$.

Discussion

Development, refining, and testing of our initial 57 items produced a smaller, 16-item EvA scale with four factors representing reasoning tendencies with respect to evidence and authority: Pro-evidence, Anti-evidence, Pro-authority, and Anti-authority. Thus, people can have separable motivations and tendencies to seek or discredit evidence and rely upon or resist authority figures as sources of advice.

Study 3: Confirmatory Factor Analysis and Convergent and Discriminant Validity

The purpose of Study 3 was to assess preregistered hypotheses regarding convergent and discriminant validity and dimensionality of the EvA scale. We hypothesized that Pro-evidence reflects effortful thinking more than numerical ability; Anti-evidence reflects distrust in science more than general pessimism; Pro-authority reflects pro-authoritarianism more than dispositional trust, Anti-authority reflects anti-authoritarianism more than dispositional distrust. We also conducted CFA to assess the dimensionality of the EvA scale on this independent sample. We additionally examined factor structure invariance between gender and education groups. Preregistration of Study 3 is available at: https://aspredicted.org/RTS_PWL.

Confirmatory Factor Analysis

Methods

We recruited 323 US adults through Prolific, an online crowdsourcing platform. Studies found that Prolific provides higher quality data compared to alternative platforms, demonstrated through better performance on attention checks, lower dishonesty, and reproducibility (Palan & Schitter, 2018; Peer et al., 2017). Seven participants who missed at least one of four attention-

checks were excluded, leaving 316 for analysis. Similar numbers of individuals with and without a college degree were recruited (college: 156; non-college: 160; demographics in Table S4).

Results

The EvA scale had an acceptable model fit in CFA on this independent sample, $\chi^2(df)$ = 209.99 (98), RMSEA = .060, SRMR = .071, CFI = .928, TLI = .912 (Table S10). No items were highly skewed to suggest removal (> ±2.0; Cassidy, 2005). All items loaded onto their corresponding factors with high standardized factor loadings (.56-.83, *p*s < .01), and all subscales had acceptable internal consistency (α = 68-.84) (Table S9). Model comparisons results using Chi-square difference tests indicated that our proposed model had a significantly better fit compared to alternatives (*p*s < .01; Table S11). These results suggested again that EvA tendencies reflect four distinct factors, consistent with our preregistered hypothesis.

Assessing Convergent and Discriminant Validity of the EvA Scale

Methods

To assess convergent and divergent validity of the EvA scale, participants completed various existing scales after the EvA, before demographic items (order of scales and items randomized): Need for Cognition Scale (Coelho et al., 2018), Distrust in Science (Nadelson et al., 2014; 12 distrust items), Right-Wing Authoritarianism (Bizumic & Duckitt, 2018), Numeracy (Weller et al., 2013), Pessimism (Scheier et al., 2012), Dispositional Trust/Distrust (Bianchi & Brockner, 2012), and Social Desirability Scale (Fischer & Fick, 1993).

Results

To demonstrate construct validity—the extent that our scale measures the construct it purports to measure (Campbell & Fiske, 1959)—we adopted the correlational approach (Hawes et al., 2012; Stöber, 2001; Watson et al., 1995), indicating convergent validity with moderate to

strong correlation coefficients (e.g., .3-.6) and discriminant validity with weaker coefficients (e.g., ≤ .2) (Anastasi & Urbina, 1997; Ward et al., 2009). We additionally conducted tests of differences in dependent correlations (Haws et al., 2012; Steiger, 1980) to compare the relative strength that EvA tendencies were related with convergent versus divergent scales.

Table 4

Convergent and Discriminant Validity of the EvA scale with Other Relevant Scales

EvA tendency	Correlation with convergent construct	Correlation with discriminant construct	Test of relative correlation strength (t-value)
Pro-evidence	.39*** (Need for Cognition)	.07 (Numeracy)	4.73***
Anti-evidence	.78*** (Distrust in Science)	–.02 (Pessimism)	14.7***
Pro-authority	.64*** (Pro-Authoritarianism)	.0.04 (Dispositional Trust)	9.02***
Anti-authority	.29*** (Anti-Authoritarianism)	.23*** (Dispositional Distrust)	0.76

Note. Entries are bivariate correlations between the EvA subscale and convergent/divergent scales. *$p < .1$; **$p < .05$; ***$p < .01$.

Convergent and divergent validity was indicated as most existing scales correlated with the respective EvA tendency as expected, with significant differences between strongly versus weakly related constructs (Table 4). Pro-evidence was more strongly correlated with Need for Cognition ($r = .39$) than Numeracy ($r = .07$), and the difference between them was significant, $t = 4.73$, $p < .01$. Anti-evidence was more correlated with Distrust in Science ($r = .78$) than Pessimism ($r = –.02$), and the difference was significant, $t = 14.7$, $p < .01$. Pro-authority was more strongly correlated with Pro-Authoritarianism ($r = .64$) than Dispositional Trust ($r = .04$), and the difference was significant, $t = 9.02$, $p < .01$. Anti-authority correlations fell just below the recommended cutoffs, with Anti-Authoritarianism being slightly less strongly related than expected for convergent validity ($r = .29$) and Dispositional Distrust being slightly more strongly related than expected for divergent validity ($r = .23$), and their difference was not significant, $t =$

0.76, p = .45. Despite this, we considered these values close enough to recommended criteria to indicate convergent and discriminant validity (Anastasi & Urbina, 1997; Ward et al., 2009).

To test our expectation that EvA tendencies do not simply reflect the desire to appear socially acceptable, but the social desirability batteries adopted for Study 3 (Strahan & Gerbasi, 1972; Fischer & Fick, 1993) unexpectedly failed to load together as a single construct (Table S17). Thus, social desirability was examined again in Study 4 with an alternative measure (Hart et al., 2015).

Assessing Measurement Invariance of the EvA Scale

To assess generalizability of the EvA, we tested for measurement invariance across education and gender groups, to ensure the scale measures the same underlying constructs with equivalent relationships across subpopulations (Brown, 2015). We conducted multiple-group CFAs in the college (n = 156) and non-college (n = 160) samples (Table S13), and examined measurement invariance in three steps: configural, metric, and scalar (Steinmetz et al., 2009; Putnick & Borstein, 2016; residual invariance discussed in supplementary materials). Configural invariance testing, indicated that the simultaneous equal form solution has an acceptable model fit, $\chi^2(196) = 322.56$, RMSEA= 0.064, SRMR= 0.071, CFI = 0.913, TLI = 0.893. This supports the same number of factors and pattern of fixed and free parameters (Steinmetz et al., 2009) between college and non-college groups (Table S8). Metric (weak) measurement invariance was supported as the equality constraints on factor loadings across education groups did not significantly degrade model fit, $\chi^2_{diff}(12) = 12.92$, p = 0.37. Scalar (strong) measurement invariance was supported as equality constraints on factor loadings and item intercepts did not significantly degrade model fit, $\chi^2_{diff}(12) = 16.58$, p = 0.17. These data suggest that the EvA generalizes as a measure of reasoning tendencies between lower and higher education groups.

Using the same steps for gender groups (male: 145; female: 162; Table S14), the initial model that tests configural invariance had an acceptable fit, $\chi^2(196) = 321.26$, RMSEA= 0.065, SRMR= 0.077, CFI = 0.92, TLI = 0.90, supporting equal factor structures between females and males (Table S15). Metric invariance was also supported, as equality constraints on factor loadings across genders did not significantly degrade the fit, $\chi^2_{diff}(12) = 15.24$, $p = 0.23$. Scalar invariance was supported, as equality constraints on factor loadings and item intercepts did not significantly degrade model fit, $\chi^2_{diff}(12) = 11.02$, $p = 0.53$. These results suggest that the EvA generalizes as a measure of reasoning tendencies between males and females.

Discussion

In Studies 2 and 3, CFA verified that the EvA scale consists of 16 items that are best explained by four underlying constructs: Pro-evidence, Anti-evidence, Pro-authority, and Anti-authority. This confirmed our preliminary findings from the EFA in Study 1. Our findings also supported convergent and divergent validity, since all four EvA subscales were more related to theoretically similar constructs (Need for Cognition, Disbelief in Science, Pro-Authoritarianism, Anti-authoritarianism) than to constructs that are less specific than what we intended to measure (Numeracy, Pessimism, Dispositional Distrust and Trust). Data also supported the idea that the EvA factor structure generalizes between lower and higher education groups and males and females.

Study 4: Criterion and Discriminant Validity of the EvA Scale

Criterion validity of a scale is demonstrated when the scale is highly correlated with the observable outcomes of external criteria, such as conceptually-related attitudes or behaviors (DeVellis, 2017; Motta et al., 2021). For instance, a measure of hockey players' aggressiveness should be highly correlated with minutes spent in the penalty box for aggression that season

(Bushman & Wells, 1998). We focused on concurrent over predictive validity (Drost 2011) because our criterion variables were measured at the same time as the EvA tendencies. Study 4 also assessed a pre-registered hypothesis about the discriminant validity of the EvA scale compared to social desirability. Preregistration of Study 4 is available at: https://aspredicted.org/9HC_B7G.

Methods

We recruited 540 US adults through the survey platform Prolific. To address a Prolific issue at the time with gender imbalances (Charalambides, 2021), we balanced recruitment between males and females. 11 participants who missed at least one of four attention checks were removed, leaving 529 for analysis (college: 266, non-college: 263; males: 255, female: 269, gender self-identifying: 5; demographics in Table S5).

Measures

Questionnaire wordings are provided in supplementary materials. All variables were constructed as the average of the constituent items, scaled to range from 0 to 1.

Susceptibility to Health Misinformation. Participants viewed eight social media posts about cancer treatments (from Scherer et al., 2021), presented in random order, in which four contained true information and four contained false information. For each, participants indicated perceived accuracy of the given information (four-point scale, "completely false" to "completely true"). Susceptibility to misinformation was measured as the average perceived accuracy of false social media posts (α = .74). We confirmed that 97.5% of participants regularly used at least one social media platform (Facebook: 68.1%, Twitter: 54.4%, Instagram: 62.8%), rendering the task externally valid.

Adherence to CDC Guidelines on COVID-19. To measure the degree to which individuals abided by the Centers for Disease Control and Prevention (CDC) guidelines on COVID-19, we used fourteen items from the CDC recommendations, such as six-foot social distancing, hand washing, and wearing a face mask (after Graupensperger et al., 2021). Participants indicated the degree to which they engaged in each activity (five-point scale, "never" to "all the time"). We adjusted items to reflect updated CDC guidelines at the time of data collection (CDC, 2022), such as changing avoiding all social gatherings to those that took place indoors. The CDC was not mentioned per se, to avoid bias. EFA revealed that these fourteen items loaded onto two factors, distancing behaviors (ten items: e.g., six-feet distancing, wearing mask) and sanitizing behaviors (three items: e.g., hand washing, disinfecting surfaces); only one item, "getting tested when feeling sick," did not meaningfully load on either (Table S18). Because sanitizing is confounded by distancing (i.e., if you distance well you need not sanitize), we used the composite of the ten distancing behaviors to indicate adherence to CDC guidelines, $\alpha = .93$.

We additionally measured confidence in the safety and efficacy of COVID-19 vaccines (Vaccine Confidence Survey Question Bank; CDC, 2021) by measuring the likelihood that respondents would recommend the COVID-19 vaccine to others, the degree that they thought the COVID-19 vaccine was safe, and how confident they were that research produced a safe and effective COVID-19 vaccine (five-point scale, "not at all" to "extremely, $\alpha = .95$).

Science Curiosity. To measure the degree to which individuals enjoy consuming scientific information (Kahan et al., 2017), for efficiency, we adopted the reduced-form science curiosity scale (Motta et al., 2021). As planned in our preregistration, we replaced an item about attending public lectures that was precluded by the pandemic with an item about conversations

about science from the original long-form scale (Kahan et al. 2017). Questions on a variety of topics (e.g., politics, religion, scientific research, celebrities) were presented, so that respondents would not infer the purpose of the questions, all on a four-point scale: following the news on new technologies ("not at all" to "very closely"), reading books on scientific research or discoveries ("none" to "more than three books"), interest in scientific research or discoveries ("not at all interested" to "very interested"), and frequency of discussing scientific research with friends, family, or co-workers ("never" to "often"), $\alpha = .74$.

Religiosity. To measure religiosity, we selected the five items from the Rohrbaugh and Jessor religiosity scale (1975) that did not contain archaic language and were applicable across religions, on a five-point scale. Participants indicated how often they attend religious services ("never" to "every week"), pray or practice religious meditation ("never" to "very often"), consider religious advice when having serious personal problems ("never" to "very often"), whether religion influences daily life ("no influence" to "a large influence") and whether they believe in the existence of God ("I don't believe" to "I am sure"), $\alpha = .93$.

Results

The relationships between the four EvA tendencies and potentially relevant attitudes and behaviors were largely consistent with our expectations about criterion validity, while illustrating the types of behaviors or attitudes that each tendency uniquely predicts. Criterion validity was assessed with ordinary least squares (OLS) with robust standard errors following our preregistered model specification (Table 5). We additionally confirmed the robustness of our findings with control variables (gender, age, education, and income), which did not alter the direction or statistical significance of the relationships (Table S19).

Table 5

Relationships between EvA Reasoning Tendencies and Criterion Behaviors or Attitudes

EvA Reasoning Tendencies	Susceptibility to health misinformation			Adherence to the CDC guide on social distancing			Confidence in COVID vaccine		
	b	95% CI	*t*	*b*	95% CI	*t*	*b*	95% CI	*t*
Pro-evidence	0.003	[–0.09; 0.10]	0.1	0.24	[0.09; 0.40]	3.0***	0.19	[0.05; 0.33]	2.6***
Anti-evidence	0.39	[0.31; 0.47]	9.6***	–0.23	[–0.35; –0.12]	–3.9***	–0.79	[–0.90; –0.68]	–14.1***
Pro-authority	0.17	[0.08; 0.26]	3.8***	–0.03	[–0.15; 0.10]	–0.4	0.37	[0.24; 0.51]	5.4***
Anti-authority	–0.05	[–0.14; 0.05]	–0.9	–0.13	[–0.27; –0.00]	–2.0**	–0.21	[–0.36; –0.07]	–2.8***
Constant	0.09	[–0.01; 0.19]	1.8*	0.71	[0.56; 0.87]	9.0***	0.81	[0.65; 0.96]	10.2***
N	529			529			529		
Adjusted R^2	.27			.08			.33		

EvA Reasoning Tendencies	Science Curiosity			Religiosity		
	b	95% CI	*t*	*b*	95% CI	*t*
Pro-evidence	0.64	[0.54; 0.74]	12.2***	0.06	[–0.09; 0.22]	0.8
Anti-evidence	–0.14	[–0.24; –0.05]	–2.9***	0.41	[0.28; 0.55]	6.1***
Pro-authority	0.06	[–0.05; 0.17]	1.2	0.21	[0.05; 0.36]	2.7***
Anti-authority	–0.04	[–0.16; 0.07]	–0.7	–0.18	[–0.34; –0.02]	–2.2**
Constant	0.09	[-0.02; 0.20]	1.6	0.20	[0.04; 0.37]	2.4**
N	529			529		
Adjusted R^2	.22			.13		

*Note*s. *b* = unstandardized regression coefficient from OLS with robust standard errors. CI = confidence interval. *t* = t-value for regression coefficient. All variables were scaled to range from 0 to 1. *$p < .1$; **$p < .05$; ***$p < .01$.

Susceptibility to Health Misinformation. People who were more Anti-evidence and more Pro-authority were more likely to perceive health misinformation as accurate; Pro-evidence and Anti-authority were unrelated. These associations mimic dynamics on social media wherein people who resist scientific evidence or follow the advice of preferred sources of information can be misinformed by weakly-supported claims by authority figures such as politicians, influencers, or celebrities (Brennen et al., 2020; Bruns et al., 2021).

COVID-19 Behaviors and Attitudes. Social distancing was more strictly followed by people who were more Pro-evidence and was less followed among those who were more Anti-evidence and Anti-authority. Thus, the CDC may operate as both a source of credible scientific evidence and an authority, supporting the utility of our four-dimensional EvA tendencies for

understanding people's varied responses. Similarly, confidence in the COVID-19 vaccine was higher for individuals who were more Pro-evidence and Pro-authority and lower for people who were more Anti-evidence or Anti-authority.

Science Curiosity. Individuals who were more Pro-evidence and those who were less Anti-evidence reported greater science curiosity, which was unrelated to Pro- or Anti-authority tendencies.

Religiosity. for individuals who were more Pro-authority or Anti-evidence and was lower for those who were more Anti-authority. This pattern reflects cases in which some highly religious people follow religious leaders whose teachings diverge from scientific evidence (Harding, 2014; O'Neill, 2021), but the pattern was not like that revealed for COVID-19 recommendations or vaccine trust—whereby people scored lower if they were either Anti-authority or Anti-evidence. This is sensible given the context of the time, in which the CDC was perceived as an authority figure, for instance by the prominence of its director, Dr. Anthony Fauci (Vlasceanu & Coman, 2022). This difference between COVID-related attitudes and religiosity highlights how EvA tendencies can characterize people's unique response to authority figures in different domains. For instance, federal health agencies promote science but can also be associated with an unevenly-trusted authority figure (Lee et al., 2016) and trusted religious authorities may give advice that is not grounded in scientific evidence (Harding, 2014; O'Neill, 2021).

Assessing Discriminant Validity of the EvA scale

We tested our preregistered hypothesis that EvA tendencies are distinct from social desirability using Impression Management items on the Balanced Inventory of Desirable Responding (Hart et al., 2015). As hypothesized, social desirability was only weakly correlated

with all four EvA tendencies, Pro-evidence (r = .17), Anti-evidence (r = .05), Pro-authority (r = –.06), Anti-authority (r = –.10), indicating discriminant validity (with correlations ≤ .2; Anastasi & Urbina, 1997; Ward et al., 2009).

Discussion

Study 4 demonstrates the criterion validity of EvA scale, highlighting its capacity to explain or predict important, real-world attitudes or behaviors that are germane to people's reliance on evidence or authority figures. This includes susceptibility to health misinformation (higher Anti-evidence and Pro-authority) and science curiosity (stronger Pro-evidence, weaker Anti-evidence). The EvA scale contributes to our understanding of individual differences in confidence in authority figures that are sometimes but not always guided by science (e.g., the CDC versus social media or religion). Interestingly, Pro-authority tendencies promote confidence in the COVID vaccine but make people more vulnerable to health misinformation on social media—because both can be considered authority figures to be heeded. Individuals who are more Anti-evidence adhered less to CDC distancing recommendations to avoid COVID-19, and were more religious. These EvA tendencies were also distinct from social desirability.

Conclusions

There are many instances in human history where evidence improved the well-being of individuals and society, such as in our opening example of measles vaccines (Rota et al., 2016). Despite this power to improve lives, some people are inclined to discredit evidence and value the opinion of preferred authority figures (Martinez-Berman et al., 2020). In our current, high-choice media environment, where anti-science sentiments exist alongside misinformation, it is important to understand when people rely upon evidence or authorities so that we can promote informed decisions and mitigate the impact of misinformation. There were no existing scales that captured

these reasoning tendencies or their underlying conceptual structure, so we developed and validated the reasoning through Evidence versus Authority (EvA) scale.

Our results suggest that people's reasoning involves separable tendencies for valuing or devaluing scientific evidence, as well as valuing or discrediting authorities. There were multiple possible underlying dimensional structures, which had never been tested, such as a single dimension from relying more on evidence to authorities or one dimension for being trusting versus untrusting for both evidence and authorities. Our rigorous testing revealed and replicated the separability of four dimensions, through exploratory and confirmatory factor analyses, on four large, independent samples. The four-factor model generalized across education and gender group. Four EvA subscales (Pro-evidence, Anti-evidence, Pro-authority, Anti-authority) can also be separated out, to efficiently meet the goals of one's research, such as to predict beliefs in false claims (e.g., reflexive open-mindedness; Pennycook & Rand, 2019) or partisan-motivated misinformation or conspiracy theories (Miller et al., 2016). This scale builds on related, existing scales that measure effortful thinking or attitudes toward hierarchical authorities, but distinctly measures tendencies to approach or avoid evidence and authorities for daily decisions. EvA tendencies are strongly correlated with conceptually similar existing scales but weakly related to more generic tendencies.

Our study has important implications for society, in which figures like politicians, celebrities, social media influencers, religious leaders, or even acquaintances can promote behaviors that defy evidence (Brennen et al., 2020; Bruns et al., 2021; Harding, 2014). We linked EvA tendencies to attitudes and behaviors in key social issues, including health misinformation, COVID-19 avoidance, science, and religion. Individuals who are more suspicious of evidence and follow authorities are more susceptible to health misinformation.

Those who trust evidence and authorities also trust the COVID vaccine more and people who discredit evidence and authorities followed CDC social distancing guidelines less. People who seek evidence or less suspicious about scientific evidence were also more curious about science. Individuals who follow authorities more or are less resistant to them were more religious. The fact that the pattern of the four EvA tendencies changes by context speaks to their utility in specifying a range of natural behavior and helping us identify and improve reasoning across essential, everyday contexts like medicine, politics, and public policy.

Our study is limited by the use of online, crowdsourced samples. Online samples are more demographically diverse than college student samples, and yield similar results to nationally-representative samples (Buhrmester et al., 2011; Coppock, 2019; Paolacci & Chandler, 2014). We improved generalizability and validity by recruiting similar numbers of individuals with higher and lower levels of education, and excluding those who indicated lapses in attention. However, our samples still overrepresented younger and more liberal people compared to the US population. We also did not yet test this scale in a non-US population, and it is possible that some of these relationships are specific to the US or at least countries that are currently enmeshed in a polarized, partisan environment. Cross-cultural research is needed.

In response to widespread concerns about the politicization of science and the spread of fabricated news (Bolsen & Druckman, 2015; Vosoughi et al., 2018), our study examined reasoning through evidence versus authority, laying the groundwork for future research on information processing, belief formation, and decision making across contexts—including in the high-choice media environments characterized by partisan polarization. The ability to assess these reasoning tendencies can facilitate instructional programs, communication strategies, and

514 interventions that improve evidence-based reasoning, tailored to individuals or populations,
515 ultimately contributing to the health of society and our democracy.

516 **References**

517 Abramowitz, A. I., & Webster, S. (2016). The rise of negative partisanship and the
518 nationalization of U.S. elections in the 21st century. *Electoral Studies*, *41*, 12–22.
519 Akobeng, A. (2005). Principles of evidence based medicine. *Archives of Disease in Childhood*,
520 *90*(8), 837–840.
521 Altemeyer, B. (1988). *Enemies of freedom: Understanding right-wing authoritarianism*. Jossey-
522 Bass.
523 Ames, S. L., Wolffsohn, J. S., & McBrien, N. A. (2005). The development of a symptom
524 questionnaire for assessing virtual reality viewing using a head-mounted display. *Optometry*
525 *and Vision Science*, 82(3), 168–176.
526 Anastasi, A., Urbina, S.,(1997). *Psychological testing* (7th ed.). NJ: Prentice Hall.
527 Baker, R., Thomas, S., Thomas, P. W., Gower, P., Santonastaso, M., & Whittlesea, A. (2010).
528 The Emotional Processing Scale: Scale refinement and abridgement (EPS-25). *Journal of*
529 *Psychosomatic Research*, *68*(1), 83–88.
530 Beavers, A. S., Lounsbury, J. W., Richards, J. K., Huck, S. W., Skolits, G. J., & Esquivel, S. L.
531 (2013). Practical Considerations for Using Exploratory Factor Analysis in Educational
532 Research. *Practical Assessment, Research & Evaluation*, *18*, 6.
533 Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*,
534 *107*(2), 238–246.

Berinsky, A. J., Margolis, M. F., Sances, M. W., & Warshaw, C. (2019). Using screeners to measure respondent attention on self-administered surveys: Which items and how many? *Political Science Research and Methods*, *9*(2), 430–437.

Bianchi, E. C., & Brockner, J. (2012). In the eyes of the beholder? The role of dispositional trust in judgments of procedural and interactional fairness. *Organizational Behavior and Human Decision Processes*, *118*(1), 46–59.

Bizumic, B., & Duckitt, J. (2018). Investigating Right Wing Authoritarianism With a Very Short Authoritarianism Scale. *Journal of Social and Political Psychology*, *6*(1), 129–150.

Bland, J. M., & Altman, D. G. (1997). Statistics notes: Cronbach's alpha. *BMJ*, *314*(7080), 572.

Bolsen, T., & Druckman, J. N. (2015). Counteracting the politicization of science. *Journal of Communication*, *65*(5), 745–769.

Bolsen, T., & Druckman, J. N. (2018). Do partisanship and politicization undermine the impact of a scientific consensus message about climate change? *Group Processes & Intergroup Relations*, *21*(3), 389–402.

Brennen, S., Simon, F. M., Howard, P. N., & Nielsen, R. K. (2020). Types, sources, and claims of COVID-19 misinformation. *Reuters Institute for the Study of Journalism*. https://reutersinstitute.politics.ox.ac.uk/types-sources-and-claims-covid-19-misinformation

Bronstein, M. V., Pennycook, G., Bear, A., Rand, D. G., & Cannon, T. D. (2019). Belief in fake news is associated with delusionality, dogmatism, religious fundamentalism, and reduced analytic thinking. *Journal of Applied Research in Memory and Cognition*, *8*(1), 108–117.

Brown, N. J. S., Furtak, E. M., Timms, M., Nagashima, S. O., & Wilson, M. (2010). The evidence-based reasoning framework: Assessing scientific reasoning. *Educational Assessment*, *15*(3–4), 123–141.

Brown, T. A. (2015). *Confirmatory factor analysis for applied research, Second Edition*. Guilford Publications.

Browne, M., Thomson, P., Rockloff, M. J., & Pennycook, G. (2015). Going against the herd: Psychological and cultural factors underlying the 'vaccination confidence gap.' *PLOS ONE*, *10*(9), e0132562.

Bruns, A., Harrington, S., & Hurcombe, E. (2021). Coronavirus Conspiracy Theories: Tracing Misinformation Trajectories from the Fringes to the Mainstream. In M. Lewis, E. Govender, & K. Holland (Eds.), *Communicating COVID-19: Interdisciplinary Perspectives* (pp. 229–249). Springer International Publishing.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, *6*(1), 3–5.

Bushman, B. J., & Wells, G. L. (1998). Trait aggressiveness and hockey penalties: Predicting hot tempers on the ice. *Journal of Applied Psychology*, *83*(6), 969–974.

Byrne, B. M. (1994). Testing for the factorial balidity, replication, and invariance of a measuring instrument: A paradigmatic application based on the Maslach Burnout Inventory. *Multivariate Behavioral Research*, *29*(3), 289–311.

Cable, D. M., & DeRue, D. S. (2002). The convergent and discriminant validity of subjective fit perceptions. *Journal of Applied Psychology*, *87*(5), 875–884.

Cacioppo, J. T., Petty, R. E., & Kao, C. F. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment*, *48*(3), 306–307.

Cassidy, D. J., Hestenes, L. L., Hegde, A., Hestenes, S., & Mims, S. (2005). Measurement of quality in preschool child care classrooms: An exploratory and confirmatory factor analysis

of the early childhood environment rating scale-revised. *Early Childhood Research Quarterly*, *20*(3), 345–360.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, *1*(2), 245–276.

CDC. (2022). *How to Protect Yourself & Others*. *Centers for Disease Control and Prevention*. https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/prevention.html

CDC. (2021). Vaccine confidence survey question bank. *Centers for Disease Control and Prevention, 24*. https://www.cdc.gov/vaccines/covid-19/vaccinate-with-confidence/rca-guide/downloads/CDC_RCA_Guide_2021_Tools_AppendixD_Surveys-508.pdf

Charalambides, N. (2021). We recently went viral on TikTok—Here's what we learned. *Prolific*. https://blog.prolific.co/we-recently-went-viral-on-tiktok-heres-what-we-learned/

Coelho, G. L. de H., H. P. Hanel, P., & J. Wolf, L. (2018). The very efficient assessment of need for cognition: developing a six-item version. *Assessment*, *27*(8), 1870–1885.

Coppock, A. (2019). Generalizing from survey experiments conducted on Mechanical Turk: A replication approach. *Political Science Research and Methods*, *7*(3), 613–628.

DeVellis, R. F. (2017). *Scale development: Theory and applications* (Vol. 26). Sage publications.

Drost, E. A. (2011). Validity and reliability in social science research. *Education Research and Perspectives*, *38*(1), 105–123.

Druckman, J. N., Peterson, E., & Slothuus, R. (2013). How elite partisan polarization affects public opinion formation. *The American Political Science Review*, *107*(1), 57–79.

Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology*, *92*(6), 1087–1101.

603 Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use
604 of exploratory factor analysis in psychological research. *Psychological Methods*, *4*(3), 272–
605 299.
606 Farias, M., Newheiser, A.-K., Kahane, G., & de Toledo, Z. (2013). Scientific faith: Belief in
607 science increases in the face of stress and existential anxiety. *Journal of Experimental Social*
608 *Psychology*, *49*(6), 1210–1213.
609 Ferguson, E., & Cox, T. (1993). Exploratory factor analysis: A users' guide. *International*
610 *Journal of Selection and Assessment*, *1*(2), 84–94.
611 Fischer, D. G., & Fick, C. (1993). Measuring social desirability: Short forms of the Marlowe-
612 Crowne social desirability scale. *Educational and Psychological Measurement*, *53*(2), 417–
613 424.
614 Francis, L. J., & Greer, J. E. (1999). Attitude toward science among secondary school pupils in
615 northern Ireland: Relationship with sex, age and religion. *Research in Science &*
616 *Technological Education*, *17*(1), 67–74.
617 Gambrill, E. (1999). Evidence-based practice: An alternative to authority-based practice.
618 *Families in Society*, *80*(4), 341–350.
619 Gorsuch, R. L., (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
620 Graupensperger, S., Lee, C. M., & Larimer, M. E. (2021). Young adults underestimate how well
621 peers adhere to COVID-19 preventive behavioral guidelines. *The Journal of Primary*
622 *Prevention*, *42*(3), 309–318.
623 Guzelian, P. S., & Guzelian, C. P. (2004). Authority-based explanation. *Science*, *303*(5663),
624 1468–1469.

Harding, D. (2014, January 12). South African preacher tells parishioners to eat grass to get closer to God. *New York Daily News*.

Hart, C. M., Ritchie, T. D., Hepper, E. G., & Gebauer, J. E. (2015). The balanced inventory of desirable responding short form (BIDR-16). *SAGE Open*, *5*(4), 2158244015621113.

Haws, K. L., Naylor, R. W., Coulter, R. A., & Bearden, W. O. (2012). Keeping it all without being buried alive: Understanding product retention tendency. *Journal of Consumer Psychology*, *22*(2), 224–236.

Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods*, *7*(2), 191–205.

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.

Huber, C. R., & Kuncel, N. R. (2016). Does College Teach Critical Thinking? A Meta-Analysis. *Review of Educational Research*, *86*(2), 431–468.

Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., & Westwood, S. J. (2019). The origins and consequences of affective polarization in the United States. *Annual Review of Political Science*, *22*(1), 129–146.

Jamieson, K. H., McNutt, M., Kiermer, V., & Sever, R. (2019). Signaling the trustworthiness of science. *Proceedings of the National Academy of Sciences*, *116*(39), 19231–19236.

Johnson, D. R., & Borden, L. A. (2012). Participants at your fingertips: Using Amazon's Mechanical Turk to increase student–faculty collaborative research. *Teaching of Psychology*, *39*(4), 245–251.

Kahan, D. M. (2015). Climate-science communication and the measurement problem. *Political Psychology*, *36*(S1), 1–43.

Kahan, D. M., Landrum, A., Carpenter, K., Helft, L., & Hall Jamieson, K. (2017). Science curiosity and political information processing. *Political Psychology*, *38*(S1), 179–199.

Lee, C., Whetten, K., Omer, S., Pan, W., & Salmon, D. (2016). Hurdles to herd immunity: Distrust of government and vaccine refusal in the US, 2002–2003. *Vaccine*, *34*(34), 3972–3978.

Lipkus, I. M., Samsa, G., & Rimer, B. K. (2001). General performance on a numeracy scale among highly educated samples. *Medical Decision Making*, *21*(1), 37–44.

Litman, L., Robinson, J., & Abberbock, T. (2017). TurkPrime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods*, *49*(2), 433–442.

Martinez-Berman, L., McCutcheon, L., & Huynh, H. P. (2020). Is the worship of celebrities associated with resistance to vaccinations? Relationships between celebrity admiration, anti-vaccination attitudes, and beliefs in conspiracy. *Psychology, Health & Medicine*, 1–10.

McDonald, R. P., & Ho, M.-H. R. (2002). Principles and practice in reporting structural equation analyses. *Psychological Methods*, *7*(1), 64–82.

McNeill, K. L., & Martin, D. M. (2011). Claims, evidence, and reasoning. *Science and Children*, *48*(8), 52–56.

Medsker, G. J., Williams, L. J., & Holahan, P. J. (1994). A review of current practices for evaluating causal models in organizational behavior and human resources management research. *Journal of Management*, *20*(2), 439–464.

Miller, J. M., Saunders, K. L., & Farhart, C. E. (2016). Conspiracy endorsement as motivated reasoning: The moderating roles of political knowledge and trust. *American Journal of Political Science*, *60*(4), 824–844.

Morof, D., Steinauer, J., Haider, S., Liu, S., Darney, P., & Barrett, G. (2012). Evaluation of the London measure of unplanned pregnancy in a United States population of women. *PloS One*, *7*(4), e35381.

Motta, M., Chapman, D., Haglin, K., & Kahan, D. (2021). Reducing the administrative demands of the science curiosity scale: A Validation Study. *International Journal of Public Opinion Research*, *33*(2), 215–233.

Nadelson, L., Jorcyk, C., Yang, D., Smith, M. J., Matson, S., Cornell, K., & Husting, V. (2014). I just don't trust them: The development and validation of an assessment instrument to measure trust in science and scientists. *School Science and Mathematics*, *114*(2), 76–86.

Nelson, R. (2019). US measles outbreak concentrated among unvaccinated children. *The Lancet Infectious Diseases*, *19*(3), 248.

Newton, C., Feeney, J., & Pennycook, G. (2021). *On the disposition to think analytically: Four distinct intuitive-analytic thinking styles*. PsyArXiv. (Working paper)

O'Neill, J. (2021, February 8). Iranian cleric says COVID vax turns people into homosexuals. *New York Post*. https://nypost.com/2021/02/08/iranian-cleric-says-covid-vax-turns-people-into-homosexuals/

Palan, S., & Schitter, C. (2018). Prolific.ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, *17*, 22–27.

Paolacci, G., & Chandler, J. (2014). Inside the Turk: Understanding Mechanical Turk as a participant pool. *Current Directions in Psychological Science*, *23*(3), 184–188.

693 Pennycook, G., & Rand, D. G. (2019). Who falls for fake news? The roles of bullshit receptivity,
694 overclaiming, familiarity, and analytic thinking. *Journal of Personality*, *88*(2), 185–200.
695 Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative
696 platforms for crowdsourcing behavioral research. *Journal of Experimental Social*
697 *Psychology*, *70*, 153–163.
698 Perry, R. T., Murray, J. S., Gacic-Dobo, M., Dabbagh, A., Mulders, M. N., Strebel, P. M., Okwo-
699 Bele, J.-M., Rota, P. A., & Goodson, J. L. (2015). Progress toward regional measles
700 elimination—worldwide, 2000–2014. *Morbidity and Mortality Weekly Report*, *64*(44),
701 1246–1251.
702 Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting:
703 The state of the art and future directions for psychological research. *Developmental Review*,
704 *41*, 71–90.
705 Pfeffer, J., & Sutton, R. I. (2006). Evidence-based management. *Harvard Business Review*,
706 *84*(1), 62–74, 133.
707 Rohrbaugh, J., & Jessor, R. (1975). Religiosity in youth: A personal control against deviant
708 behavior. *Journal of Personality*, *43*(1), 136–155.
709 Rota, P. A., Moss, W. J., Takeda, M., de Swart, R. L., Thompson, K. M., & Goodson, J. L.
710 (2016). Measles. *Nature Reviews Disease Primers*, *2*(1), 1–16.
711 Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical*
712 *Software*, *48*, 1–36.
713 Rynes, S. L., Colbert, A. E., & O'Boyle, E. H. (2018). When the "best available evidence"
714 doesn't win: How doubts about science and scientists threaten the future of evidence-based
715 management: *Journal of Management*.

Sanderson, I. (2002). Evaluation, policy learning and evidence-based policy making. *Public Administration*, *80*(1), 1–22.

Scheier, M. F., Carver, C. S., & Bridges, M. W. (1994). Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): A reevaluation of the Life Orientation Test. *Journal of Personality and Social Psychology*, *67*(6), 1063–1078.

Scherer, L. D., McPhetres, J., Pennycook, G., Kempe, A., Allen, L. A., Knoepke, C. E., Tate, C. E., & Matlock, D. D. (2021). Who is susceptible to online health misinformation? A test of four psychosocial hypotheses. *Health Psychology: Official Journal of the Division of Health Psychology, American Psychological Association*, *40*(4), 274–284.

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research*, *8*(2), 23–74.

Schommer, M. (1998). The influence of age and education on epistemological beliefs. *British Journal of Educational Psychology*, *68*(4), 551–562.

Schraw, G., Bendixen, L. D., & Dunkle, M. E. (2002). Development and validation of the Epistemic Belief Inventory (EBI). In *Personal epistemology: The psychology of beliefs about knowledge and knowing* (pp. 261–275). Lawrence Erlbaum Associates Publishers.

Shearman, S. M., & Levine, T. R. (2006). Dogmatism updated: A scale revision and validation. *Communication Quarterly*, *54*(3), 275–291.

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, *87*(2), 245–251.

737 Steinmetz, H., Schmidt, P., Tina-Booh, A., Wieczorek, S., & Schwartz, S. H. (2009). Testing
738 measurement invariance using multigroup CFA: Differences between educational groups in
739 human values measurement. *Quality & Quantity*, *43*(4), 599.
740 Stöber, J. (2001). The social desirability scale-17 (sds-17): Convergent validity, discriminant
741 validity, and relationship with age. *European Journal of Psychological Assessment*, *17*(3),
742 222–232.
743 Streiner, D. L. author, Norman, G. R., Cairney, J., Streiner, D. L., Norman, G. R., & Cairney, J.
744 (2015). *Health measurement scales: A practical guide to their development and use* (Fifth
745 edition.). Oxford University Press.
746 Strahan, R., & Gerbasi, K. C. (1972). Short, homogeneous versions of the Marlowe-Crowne
747 Social Desirability Scale. *Journal of Clinical Psychology*, *28*(2), 191–193.
748 Svedholm-Häkkinen, A. M., & Lindeman, M. (2017). Actively open-minded thinking:
749 Development of a shortened scale and disentangling attitudes towards knowledge and
750 people. *Thinking & Reasoning*, *24*(1), 21–40.
751 Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed.). New York:
752 Harper & Row.
753 Tanne, J. H. (2019). US county bars unvaccinated children from public spaces amid measles
754 emergency. *BMJ : British Medical Journal*, *364*.
755 Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of*
756 *Medical Education*, *2*, 53–55.
757 van Prooijen, J.-W. (2017). Why Education Predicts Decreased Belief in Conspiracy Theories.
758 *Applied Cognitive Psychology*, *31*(1), 50–58.

759 Vlasceanu, M., & Coman, A. (2022). The impact of information sources on COVID-19
760 knowledge accumulation and vaccination intention. *International Journal of Data Science*
761 *and Analytics*, *13*(4), 287–298.
762 Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*,
763 *359*(6380), 1146–1151.
764 Waldinger, M. D. (2004). Lifelong premature ejaculation: From authority-based to evidence-
765 based medicine. *BJU International*, *93*(2), 201–207.
766 Ward, C., Fischer, R., Zaid Lam, F. S., & Hall, L. (2009). The Convergent, Discriminant, and
767 Incremental Validity of Scores on a Self-Report Measure of Cultural Intelligence.
768 *Educational and Psychological Measurement*, *69*(1), 85–105.
769 Watson, D., Weber, K., Assenheimer, J. S., Clark, L. A., Strauss, M. E., & McCormick, R. A.
770 (1995). Testing a tripartite model: Evaluating the convergent and discriminant validity of
771 anxiety and depression symptom scales. *Journal of Abnormal Psychology*, *104*(1), 3–14.
772 Weller, J. A., Dieckmann, N. F., Tusler, M., Mertz, C. K., Burns, W. J., & Peters, E. (2013).
773 Development and testing of an abbreviated numeracy scale: A Rasch analysis approach.
774 *Journal of Behavioral Decision Making*, *26*(2), 198–212.
775 Wombwell, E., Fangman, M. T., Yoder, A. K., & Spero, D. L. (2015). Religious barriers to
776 measles vaccination. *Journal of Community Health*, *40*(3), 597–604.
777 Worthington, R. L., & Whittaker, T. A. (2006). Scale development research: A content analysis
778 and recommendations for nest practices. *The Counseling Psychologist*, *34*(6), 806–838.

Supplementary Materials

for

The Reasoning through Evidence versus Authority (EvA) Scale: Scale Development and Validation

Initial Scale Development and Pilot Study

To ensure item clarity and a fit between intended and inferred constructs, the full set of initial 57 items were tested using individual, in-person interviews with Introductory Psychology Subject Pool participants (n = 5) and adults without a college education that we recruited through flyers and craigslist advertisements (n = 4). The goal of the interviews was to ensure that questions were written in a manner participants could easily understand.

The first round of interviews was conducted on undergraduate students (n = 5), using the U-M Intro Psychology Subject Pool (August 11, 2019 ~ August 15, 2019). To assess whether non-student adults without a college degree comprehend the items in accordance with our intentions, non-student participants without college education (n = 4) were invited for a second round of interviews to provide feedback on the revised questions (October 16, 2019 ~ October 28, 2019). In the second round, participants were recruited to interview regarding the scale development by flyers posted in nearby business locations and by an ad posted on Craigslist. Both forms of advertisement used the same text asking for adults who do not have a college education to participate in our research. Participants were compensated $10 for 45 minutes of participation at the lab at the university.

Each interview was conducted by having participants complete the survey while talking through their reasoning as they answered each question. With the experimenter present, for each item participants assessed: What is the item asking you?; Is there anything confusing or ambiguous about this item?; Is there an answer choice that accurately reflects how you would like to respond? For questions that took a long time to reason through, or otherwise took a lot of effort to understand or a lot of debate back and forth between various interpretations, the participant was asked further questions about semantic definitions (i.e. What do you think of when you think of authority? What kinds of politicians are you thinking about in this question?). Later, participants described to the experimenter how well they believed the scale captured tendencies to seek or avoid evidence or authorities and suggested improvements.

It was found that many people shared similar conceptions of these terms—authority tended to refer to parents, police, and doctors, and politicians were often discussed at a national level, rather than a state or local level. Items were edited when there was non-uniformity in semantic understandings or when the wording of the question was unnecessarily difficult. The items shown in Table S1 had outstanding comments and criticisms. Other items were interpreted by interviewees as we intended, specifically whether each item pertains to individual tendencies to rely on or suspect evidence and/or authorities.

Table S1

Items Revised through Pilot Study

Item	Original item wording	Interview feedback	Revised item wording
20	I don't think the media are knowledgeable enough about the facts to rely on them for answers.	Interviewees felt that 'media' was too vague. Social media sources like facebook, twitter? Cable news? "Mainstream media" or "news media" or "the major media sources" might make this item clearer	Hosts of major television news shows do not know enough to be reliable sources of information.

25	I generally ignore scientists when they are talking about their research.	Some felt that this was less about the scientists than about the research, despite the wording seeming to focus on the scientists. Something that singles out and dismisses "scientists" would avoid this.	Scientists' research doesn't matter in the real world.
33	Before I vote on an issue in my state or city, I try to look up the details so that I vote correctly.	One thought this item was asking about whether you look up details about voting procedures ("When, where, how do I vote correctly?); another disagreed with the item because of lack of interest in politics/voting.	Before I vote on an issue in my state or city, I try to look up the ballot items so that I vote correctly.
40	In general, you should evaluate the accuracy of information in a textbook.	Mixed views on this item. The impression was that participants recognized that this is a pro-evidence item, but most wouldn't agree/strongly agree even if they were strongly pro-evidence people, because they wouldn't take the time to do fact-check textbooks or because they view textbooks as classic/time-honored sources (e.g. a college calculus textbook) not worth checking.	In general, you should consider whether the information in your textbook is accurate.

Data Collection and Descriptive Statistics

Study 1

Study 1 was conducted via the survey platform CloudResearch between March 13, 2021 and March 21, 2021. To collect similar number of respondents with and without college education, we used CouldResearch's prescreening data to recruit half of the participants without college degree, and another half with a college degree. To ensure response quality, approval rate was set at 95% and the number of approved HIT's was set to be greater than or equal to 100.

Table S2

Distribution of Demographic Variables in Study 1 (n = 549)

Variable	Distribution
Gender	Male (1) = 46.4%; Female (2) = 52.6%; Prefer to self-identify (3) = 0.9%
Race	White (1) = 83.1%; Black or African American (2) = 6.9%; American Indian or Alaska Native (3) = 0.5%; Asian (4) = 6.7%; Native Hawaiian or other Pacific Islander (5) = 0.0%; Other (e.g., mixed) (6) = 2.7%
Age	Age 18-24 = 7.3%; Age 25-34 = 24.4%; Age 35-44 = 27.7%; Age 45-54 = 18.0%; Age 55-64 = 14.2%; Age 65 or older = 8.4%
Education	No high school diploma (1) = 1.1%; High school diploma (2) = 42.4%; Some college, no degree (3) = 3.8%; Associate degree (4) = 8.4%; Bachelor's degree (5) = 26.8%; Master's degree (6) = 13.7%; Professional degree (7) = 2.6%; Doctorate degree (8) = 1.3%
Partisan Identity	Strong Democrat (1) = 19.5 %; Weak Democrat (2) = 20.4%; Democratic leaner (3) = 9.5%; Independent (4) = 14.8%; Republican leaner (5) = 7.1%; Weak Republican (6) = 15.3%; Strong Republican (7) = 13.5%
Ideology	Very liberal (1) = 12.4%; Liberal (2) = 18.0 %; Slightly liberal (3) = 13.3% ; Moderate (4) = 25.1%; Slightly conservative (5) = 9.3%; Conservative (6) = 15.5%; Very conservative (7) = 6.4%

Study 2

Study 2 was conducted via the survey platform CloudResearch between November 26, 2019 and November 27, 2019. Participants were conducted with the same set of targeting and quality controls as Study 1 (targeting by education, approval rate, number of approved HITs). Data collection for Study 2 preceded (November 2019) that of Study 1 (March 2021) for the following reason. Our earlier interpretation of EFA on Study 2 data (n=189) suggested a three-factor EvA construct, so we proceeded with follow-up data collection to validate 12 candidate items in 2020. However, through a review process, we later learned that the sample size of 189 was too small relative to the number of our initial items (58 items). To ensure reliable EFA, we decided to conduct the survey for Study 1 on a larger sample. This time, we determined the sample size on the basis of the guidelines on the minimum ratios of participants to items (5:1 or 10:1) for exploratory factor analysis (Gorsuch, 1983; Worthington & Wittaker, 2006). Given that we were at the early stage of scale development, we targeted a participant-item ratio of 10:1 and aimed to recruit 580 participants in Study 1, conducted in March 2021. Because the data collected in November 2019 contained all 58 items, the dataset was still useful as an independent sample to confirm the reliability of the CFA results, thus was included in our paper as Study 2.

Table S3

Distribution of Demographic Variables in Study 2 (n = 198)

Variable	Distribution
Gender	Male (1) = 51.9%; Female (2) = 48.1%
Race/Ethnicity	White/Caucasian (1) = 75.1%; African American (2) = 9.0%; Latino (3) = 8.5%; East Asian (4) = 3.70%; Native/Alaskan (5) = 1.1%; Other (6) = 2.6%
Age	Age 18-24 = 9.0%; Age 25-34 = 45.0%; Age 35-44 = 23.8%; Age 45-54 = 10.1%; Age 55-64 = 10.6%; Age 65 or older = 1.6%
Education	No high school diploma (1) = 2.1%; High school diploma (2) = 41.8%; Some college (3) = 9.5%; College degree (4) = 36.5%; Some post-graduate work (5) = 1.6%; Post-graduate degree (6) = 8.5%
Income	Less than $5,000 (1) = 5.8%; $5,000 to $6,999 (2) = 1.1%; $7,000 to $7,499 (3) = 1.1%; $7,500 to $9,999 (4) = 1.6%; $10,000 to $12,499 (5) = 2.1%; $12,500 to $14,999 (6) = 2.6%;$15,000 to $19,999 (7) = 5.8%; $20,000 to $24,999 (8) = 9.5%; $25,000 to $29,999 (9) = 7.4%; $30,000 to $34,999 (10) = 6.3%; $35,000 to $39,999 (11) = 6.3%; $40,000 to $49,999 (12) = 6.9%; $50,000 to $59,999 (13) = 13.2%; $60,000 to $74,999 (14) = 6.3%; $75,000 to $84,999 (15) = 6.3%; $85,000 to $99,999 (16) = 4.8%; $100,000 to $124,999 (17) = 5.8%; $125,000 to $149,999 (18) = 2.1%; $150,000 to $174,999 (19) = 1.6%; $175,000 or more (20) = 3.2%
Social Ideology	Very Liberal (1) = 16.4%; Liberal (2) = 27.5%; Moderate (3) = 28.0%; Conservative (4) = 21.2%; Very conservative (5) = 6.9%
Economic Ideology	Very Liberal (1) = 13.3%; Liberal (2) = 21.8%; Moderate (3) = 27.1%; Conservative (4) = 31.4%; Very conservative (5) = 6.4%

Study 3

Study 3 was conducted via the survey platform Prolific between June 18, 2021 and June 19, 2021. Participants were conducted via the online crowdsourcing platform Prolific, with a target to recruit an equal number of individuals with and without a college degree.

Table S4

Distribution of Demographic Variables in Study 3 (n = 316)

Variable	Distribution
Gender	Male (1) = 51.3%; Female (2) = 45.9%; Prefer to self-identify (3) = 2.8%
Race	White (1) =65.5%; Black or African American (2) = 14.2%; American Indian or Alaska Native (3) = 1.6%; Asian (4) = 12.0%; Native Hawaiian or other Pacific Islander (5) = 0.6%; Other (e.g., mixed) (6) = 6.0%
Age	Age 18-24 = 37.0%; Age 25-34 = 38.9%; Age 35-44 = 14.6%; Age 45-54 = 5.7%; Age 55-64 = 2.5%; Age 65 or older = 1.3%
Education	No high school diploma (1) = 2.5%; High school diploma (2) = 19.6%; Some college, no degree (3) = 26.9%; Associate degree (4) = 1.6%; Bachelor's degree (5) = 28.2%; Master's degree (6) = 18.0%; Professional degree (7) = 1.3%; Doctorate degree (8) = 1.9%
Religion	Protestant (1) = 13.9%; Roman Catholic (2) = 24.1%; Orthodox Christian (3) = 3.8%; Mormon (4) = 0.9%; Jewish (5) = 0.9%; Muslim (6) = 1.9%; Buddhist (7) = 1.9%; Hindu (8) = 1.6%; Atheist (9) = 15.5%; Agnostic (10) = 14.6%; Other (11) = 7.3%; Nothing in particular (12) = 13.6%
Partisan Identity	Strong Democrat (1) = 29.1 %; Weak Democrat (2) = 24.1%; Democratic leaner (3) = 13.6%; Independent (4) = 14.2%; Republican leaner (5) = 3.8%; Weak Republican (6) = 7.3%; Strong Republican (7) = 7.9%
Ideology	Very liberal (1) = 20.6%; Liberal (2) = 27.6%; Slightly liberal (3) = 12.4% ; Moderate (4) = 18.4%; Slightly conservative (5) = 5.4%; Conservative (6) = 11.7%; Very conservative (7) = 3.8%

Study 4

Study 4 was conducted via the survey platform Prolific between January 14, 2022 and January 15, 2022. Participants were conducted via the online crowdsourcing platform Prolific, with a target to recruit an equal number of individuals with and without a college degree. To address the potential issue with gender imbalance, the recruitment was balanced across gender groups.

Table S5

Distribution of Demographic Variables in Study 4 (n = 529)

Variable	Distribution
Gender	Male (1) = 48.2%; Female (2) = 50.9%; Prefer to self-identify (3) = 0.9%
Race	White (1) = 82.2%; Black or African American (2) = 7.2%; American Indian or Alaska Native (3) = 0.4%; Asian (4) = 5.9%; Native Hawaiian or other Pacific Islander (5) = 0.2%; Other (e.g., mixed) (6) = 4.2%
Age	Age 18-24 = 19.5%; Age 25-34 = 30.8%; Age 35-44 = 20.0%; Age 45-54 = 12.5%; Age 55-64 = 10.8%; Age 65 or older = 6.4%
Education	No high school diploma (1) = 0.6%; High school diploma (2) = 13.2%; Some college, no degree (3) = 26.1%; Associate degree (4) = 9.8%; Bachelor's degree (5) = 35.0%; Master's degree (6) = 12.1%; Professional degree (7) = 2.3%; Doctorate degree (8) = 0.9%
Income	Less than $10,000 (1) = 4.2%; $10,000 to $19,999 (2) = 6.8%; $20,000 to $29,999 (3) = 10.4%; $30,000 to $39,999 (4) = 9.1%; $40,000 to $49,999 (5) = 7.0%; $50,000 to $59,999 (6) = 10.8%; $60,000 to $69,999 (7) = 7.6%; $70,000 to $79,999 (8) = 9.6%;

	$80,000 to $89,999 (9) = 5.5%; $90,000 to $99,999 (10) = 4.5%; $10,0,000 to $149,999 (11) = 14.6%; More than $150,000 (12) = 10.0%
Religion	Protestant (1) = 16.8%; Roman Catholic (2) = 14.4%; Orthodox Christian (3) = 2.1%; Mormon (4) = 0.2%; Jewish (5) = 3.4%; Muslim (6) = 1.1%; Buddhist (7) = 1.1%; Hindu (8) = 0.6%; Atheist (9) = 16.3%; Agnostic (10) = 20.0%; Other (11) = 7.8%; Nothing in particular (12) = 16.3%
Partisan Identity	Strong Democrat (1) = 27.7 %; Weak Democrat (2) = 22.5%; Democratic leaner (3) = 15.7%; Independent (4) = 12.5%; Republican leaner (5) = 5.7%; Weak Republican (6) 9.8%; Strong Republican (7) = 6.1%
Ideology	Very liberal (1) = 19.0%; Liberal (2) = 26.2%; Slightly liberal (3) = 15.4% ; Moderate (4) = 17.8%; Slightly conservative (5) = 8.0%; Conservative (6) = 9.9%; Very conservative (7) = 3.8%
Social media usage	Facebook (1) = 68.0%; Twitter (2) = 53.5%; Reddit (3) = 29.1%; Instagram (4) = 62.7%; YouTube (5) = 77.0%; Do not use any kind of social media (9) = 2.4% *Note:* Sum of percentages exceed 100%, because multiple choices were allowed (i.e., "check all that apply")

Study 1 Results

Additional Details for Exploratory Factor Analysis

We chose common factors analysis (FA) rather than principal components analysis (PCA), as FA is more suitable to understand the latent constructs that explain shared variance among items in scale development (Fabrigar et al., 1999). In Study 1, for exploratory factor analysis, we used an oblique rotation for factor rotation. Fabrigar et al. (1999) suggest that using oblique rotation reduces the number of cross-factor loadings, producing superior simple structure ("cleaner" solutions). Both Fabrigar et al. (1999) and Marcus et al. (2006) recommend using oblique rotation because using orthogonal rotation will forfeit any knowledge of existing correlations among factors. Brown (2015) says that oblique rotation is preferred in most cases because it provides a more realistic representation of how factors are interrelated. We used maximum Likelihood as the model-fitting procedure, following Fabrigar et al. (1999), Cudeck & O'Dell (1994), and Brown (2015), who suggested using ML as the model-fitting procedure because it provides a wide range of fit indices and allows computation of correlations among factors, thus preferred over principal factors procedure.

Skewness (mean = –0.22, range = –1.35-1.42) and kurtosis per item (mean = 3.13, range = 1.73-5.58) suggested the appropriateness of the maximum likelihood (ML) factor extraction procedure (Baker et al., 2010; Fabrigar et al., 1999; normality violated if skewness > 2, kurtosis > 7). Worthington and Whittaker (2006) recommend that EFA be followed by examining the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy. The KMO test examines whether the associations among items can be accounted for by a smaller set of factors (Ferguson & Cox, 1993), thus indicating whether the data is adequate for examining meaningful factor structures, rather than chance correlations among a small subset of items (Worthington & Whittaker, 2006). KMO values of .60 and higher are recommended for reliable factor analysis (Beavers et al. 2013; Tabachnick & Fidell, 2001). Our KMO was .91, which suggested an adequate sample size for EFA on 57 items. Because it is advised not to include additional scales at the early stage of scale development, especially when there are a high number of initial items compared to the final scale

(Worthington & Whittaker, 2006), this study did not contain any other scales (e.g., convergent or discriminant). No items had to be removed due to highly skewed distributions (> ±2.0; Cassidy et al., 2005).

Table S6 shows the list of questionnaire items that were included in Study 1 for exploratory factor analysis. Hypothesized dimensions behind the construction of each item are indicated as: pro-evidence (E+), anti-evidence (E-), pro-authority (A+), anti-authority (A-). The acronyms for the original scale from which the items were adopted are: Schommer Epistemological Questionnaire (SEQ, Schommer, 1998; 63-item version), Epistemic Beliefs Inventory (EBI, Schraw et al., 2004), the Updated Dogmatism Scale (UDS, Shearman & Levine, 2006), and Attitudes Toward Science Scale (ATSS, Francis & Greer, 1999). Bold-faced items were selected for the final EvA scale.

Table S6

EvA Items and Exploratory Factor Analysis Results: Study 1 (n = 549)

			Factors								
Item	Statement	Expected Trait	1	2	3	4	5	6	Skew	Kurt	Comm
57	**Scientific evidence is overrated; there are often better ways to understand the world.**	**E-**	0.84						0.80	2.91	1.1
25	Scientists' research doesn't matter in the real world.	E-	0.78						1.42	5.06	1.1
50	**Even if scientific studies are done carefully and transparently, I still don't really believe them.**	**E-**	0.76						0.78	3.08	1.1
48	**People make too much of scientific studies in the news when I know that the research is biased anyway.**	**E-**	0.62						0.26	2.13	1.5
47	**When new evidence reverses a previous scientific theory, I just stop paying attention to it and make my own decisions.**	**E-**	0.57						0.36	2.18	1.2
7	My behavior is usually dictated by my religious values.	A	0.53	0.31	0.39				0.45	1.73	2.7
46	When scientists change their minds, I stop trusting their research on what we are supposed to eat to be healthy.	E-	0.51						0.31	2.33	1.6
52	I think scientific data is too hard to understand, so I generally ignore it.	E-	0.49	-0.3					1.03	3.63	2
53	When it comes to controversial issues in society, I don't think "the data" can tell us much.	E-	0.46						0.52	2.55	1.6
49	People can talk about data, but I think that my intuitions are a better guide for my decisions.	E-	0.46						0.28	2.31	1.8
1	I pay close attention to what my religious leader tells me I should do.	A	0.39		0.39				0.50	1.87	3.4
8	I assume my doctors know what they're talking about, so I follow their recommendations.	A	-0.32				0.32	-0.31	-1.11	4.74	4.2
42	It is usually wise to seek out evidence and research before making decisions. (adapted from UDS)	E+	-0.37	0.3		0.31			-1.03	4.63	3.1
39	Money spent on science is well worth spending. (adapted from ATSS)	E+	-0.83						-0.90	3.68	1.3
38	Science is very important for the country's development. (adapted from ATSS)	E+	-0.94				0.31		-1.35	5.26	1.3
32	**When I hear a news story reporting research about health, I want to look up the study they are referring to.**	**E+**		0.85					-0.84	3.45	1.2

34	**I carefully examine research on important issues to make sure it is valid and unbiased.**	**E+**	0.84			-1.09	4.78	1.1
35	When I hear about new research, I look into who funded it to be sure it is unbiased.	E+	0.71			-0.63	2.69	1.3
41	**When debating an important issue, I try to fact-check things that people state as statistics.**	**E+**	0.68			-0.90	4.17	1.1
27	**When someone makes a statement that sounds like a fact, I want to know the evidence behind it.**	**E+**	0.67			-1.02	4.47	1.2
28	When someone cites a statistic, I want to know where they got it from.	E+	0.61			-1.17	4.66	1.5
30	I pay attention to science news and try to follow the latest findings.	E+	0.61			-0.67	3.04	2.5
33	Before I vote on an issue in my state or city, I try to look up the ballot items so that I vote correctly.	E+	0.57			-1.21	4.50	1.1
37	When my doctor tells me about a new treatment, I like to find out about any research on the treatment.	E+	0.57			-1.24	5.58	1.3
26	When I hear a news story about health, I wonder if there is really good evidence behind the assertion.	E+	0.42			-0.87	4.07	1.9
36	I think news reports about science should include more information so that we can evaluate the strength of the evidence.	E+	0.39		0.36	-0.69	3.73	2.6
40	In general, you should consider whether the information in your textbook is accurate. (adapted from SEQ)	E+	0.3			-0.77	3.47	3.4
51	I am not interested in looking into the details when I hear the results of a new study.	E-	-0.57			0.82	3.02	1.5
5	I respect law enforcement, like police officers.	A		0.77		-1.04	3.46	1.1
12	People should always respect authority. (adapted from UDS)	A		0.69		-0.30	2.36	1.3
3	I generally follow my parents' advice.	A		0.42		-0.59	2.87	2
13	People who are in a position of authority have the right to tell others what to do. (adapted from UDS)	A		0.35		-0.26	2.31	2.2
15	I wonder how much my teachers really knew. (adapted from SEQ)	A-		-0.30	0.33	-0.25	2.24	2.4
18	I am doubtful that my teachers really understood what they were teaching me.	A-		-0.37		0.48	2.57	2.8
22	I assume that people in positions of power are corrupt.	A-		-0.37	0.43	-0.07	2.44	2.3
24	From my perspective, people in positions of authority should generally not be trusted.	A-		-0.54		0.36	2.85	1.8
14	Children should be allowed to question their parents' authority. (adapted from EBI)	A-		-0.56		-0.11	2.23	1.6
17	I'm the type of person who questions authority. (adapted from UDS)	A-		-0.61		-0.25	2.40	1.5
29	**I am concerned that news reports are based on people's opinions rather than actual evidence.**	**A-**			0.66	-0.73	2.79	1.3
21	**Government officials often say things that are untrue in their public statements.**	**A-**			0.66	-0.69	3.42	1.1
19	**Hosts of major television news shows do not know enough to be reliable sources of information.**	**A-**			0.61	-0.35	2.47	1.1
23	**People who are telling us how to act don't always have an incentive to tell the truth.**	**A-**			0.49	-0.56	2.96	1.4
16	Even advice from experts should be questioned. (adapted from SEQ)	A-			0.41	-0.75	3.63	2.1
20	Just because people are older or more experienced does not mean their claims are necessarily correct.	A-			0.32	-1.16	4.88	2.7

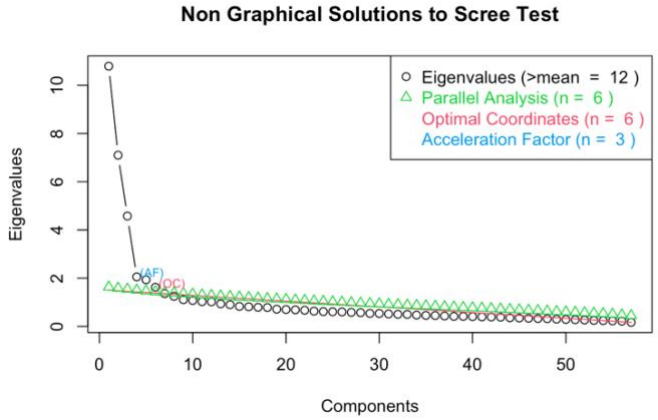
2	**When I have to vote, I see what my politician says and follow their lead.**	**A**	0.64		0.24	2.23	1.2
6	**I often make changes to my diet based on what my friends tell me is more healthy.**	**A**	0.55		0.80	3.02	1.5
9	**When I think a politician has a confident, assertive personality, I naturally like them and vote for them.**	**A**	0.54		0.18	2.40	1.3
4	**I assume that when my favorite blogger or social media personality gives advice, they know what they are talking about.**	**A**	0.51		0.46	2.43	1.4
11	Sometimes you just have to accept the teacher's answer even though you don't understand it. (adapted from SEQ)	A	0.35		-0.23	2.17	2.2
10	People who challenge authority are overconfident. (adapted from SEQ)	A	0.31		0.25	2.44	2.4
44	I am wary of medical procedures that interfere with my body's natural processes.	E-		0.79	-0.64	2.63	1.2
45	I am hesitant to take prescription medicines because they seem like chemicals I am putting into my body.	E-		0.76	0.01	1.78	1.1
43	I believe that things that are natural are always better for you.	E-		0.63	-0.57	2.64	1.6
54	I avoid GMOs and pesticides, no matter what the evidence says.	E-		0.53	0.11	2.01	1.7
56	When choosing between products, I don't spend much time comparing the specifications.	E-			0.94	3.67	2.9
55	I rarely check the nutrition facts or ingredient list on food labels.	E-			0.71	2.35	3.8
31	I am more likely to avoid a risk when I learn about the statistics rather than personal stories and anecdotes.	E+			-0.55	3.17	2.9

Note. Factor loadings smaller than .3 are not displayed. The bolded items were retained for the EvA scale. *Skew* = skewness; *Kurt* = Kurtosis; *Comm* = Communality.

Regarding assessing the number of factors, while the Kaiser test (Kaiser 1960) is a widely used criteria (finding number of factors through the number of eigenvalues greater than 1), we considered a scree plot and a parallel analysis because the eigenvalue rule can be too generous as a basis of retaining factors (DeVellis, 2006). The scree plot and parallel analysis both suggested six factors (Figure S1; Cattell, 1966; Hayton et al., 2004).

Figure S1

Parallel Analysis of Study 1



Following recommended item deletion criteria (Baker et al., 2010; Haws et al., 2012; Svedholm-Häkkinen & Lindeman, 2017; Worthington & Whittaker, 2006), eight items that cross-loaded on more than one factor > .30 (Items 1, 7, 8, 22, 36, 38, 42, 52) were dropped along with ten items whose highest loading was < .40 (Items 10, 11, 13, 15, 18, 20, 31, 40, 55, 56). No factors needed to be removed due to having fewer than three items (Baker et al., 2010; Brown 2015; Tabachnick & Fidell, 2001), with at least four items loading > .40 per factor. In terms of item communality, items with low communalities (< .40) are not highly correlated with one or more of the factors in the solution and should be dropped (Tabachnick & Fidell, 2001). Communalities of remaining items ranged between 1.0 to 2.5, thus none were dropped on the basis of communalities. Table S6 presents the full EFA results, including factor loadings, kurtosis, skewness, and communalities.

Item Retention Decisions

Because some of the factors contained a relatively large number of items, the last step of EFA was to shorten the scale, where researchers often aim for a balanced scale with similar number of items per factor (Baker et al., 2010; Worthington & Whittaker, 2006). Optimizing the scale length is recommended for the efficiency of the scale (e.g., respondent fatigue) despite its tradeoff of sacrificing a certain degree of internal consistency (Worthington & Whittaker, 2006) Since EFA is "a combination of empirical and subjective approaches to data analysis" with the goal of arriving at a solution that makes sense (Worthington & Whittaker, 2006), we actively employed both empirical and substantive rationales in item selection decisions. To have a balanced number of items across factors, because Factors 5 and 6 contained four items, we aimed to select four items for four other dimensions that contained more than four items. Following the literature (Brown, 2015; Worthington & Whittaker, 2006), we deleted items that (a) have the lowest factor loadings, (b) have the highest cross-loadings, (c) contribute to internal consistency the least, and (d) have low conceptual consistency with other items. Among the four criteria, there was no item to drop based on cross-loadings because cross-loading items were dropped in the previous step. Employing a mix of criteria (a), (c), and (d), we made the following decisions.

Table S7

Change in Reliability after Removal of Each Item

Factor 1	Alpha if item deleted	**Factor 2**	Alpha if item deleted	**Factor 3**	Alpha if item deleted	Factor 4	Alpha if item deleted
Item 57	.83	Item 32	.79	Item 5	.70	Item 29	.68
Item 25 (-)	**.85**	Item 34	.78	Item 12	.69	Item 21	.68
Item 50	.83	**Item 35**	**.82**	**Item 24 (-)**	**.78**	Item 19	.70
Item 48	.84	Item 41	.81	Item14 (-)	.76	Item23	.73
Item 47	.85	Item 27	.80	Item17 (-)	.73	**Item 16**	**.74**
5-item Alpha	*.87*	*5-item Alpha*	*.83*	*5-item Alpha*	*.78*	*5-item Alpha*	*.75*

For Factor 1 (Anti-evidence), the four lowest loading items (Items 46, 49, 53, 39) were removed. Compared to the retained items that tap onto individuals' general tendency of how they approach "scientific evidence" (Items 50, 25, 57, 48, 47), the removed items were relevant to "data," "healthy eating," or "spending on science," thus our decision to remove them was supported in terms of conceptual consistency. Although Item 39 strongly loaded on Factor 1 with a negative factor loading, we did not retain it because its focus (spending on science) was quite distant from the overall substantive meaning of this factor. Among the five retained items, we

employed the criterion (c), removing item that contributes the least to internal consistency, which is indicated by the expected increase in Cronbach's alpha for the subscale if the item is removed (Gliem & Gliem, 2003; Raubenheimer, 2004). The item analysis in Table S7 suggested that the deletion of either Item 25 or Item 47 would deteriorate internal reliability the least. Given the concerns that the item deletion decision solely based on alpha could be imperfect (Raykov, 2008), we additionally considered (d) conceptual consistency, and decided to drop Item 25, because it refers to "scientists" whereas all other remaining items are about "scientific evidence," "scientific studies," or "evidence on scientific theory." Another conceptual reason behind this decision was that Item 25 was the only reverse-coded item that we originally expected to load on the Anti-authority factor because it was intended to capture individuals' resistance toward "scientists" as authority figures. Based on these considerations, Items 50, 57, 48, and 47 were chosen for Factor 1.

For Factor 2 (Pro-evidence), the six lowest loading items (Items 28, 30, 33, 37, 26, 51) were removed. The removed items were conceptually distinct from retained items that were closely related to the general attitude toward evidence ("examine research for validity and unbiasedness," "check evidence behind claims," "look up study behind a news story"), whereas three of the deleted items (Items 33, 37, 51) were relevant to specific topics such as "medical treatment" or "voting," and one of them (Item 51) was a reverse-coded item originally intended to load onto as the Anti-evidence dimension. Because it is recommended to avoid redundancy among the retained items (Aluja et al., 2006), dropping Item 26 was reasonable because its content overlapped with one of the retained items, Item 32 (both items were about the tendency to check evidence behind a health news story). We also took notice of the redundancy between Item 28 and Item 41, which were conceptually identical, both asking about the tendency to check evidence behind other people's use of statistics. Between the two items, we removed Item 28 that had smaller loading. We also removed Item 30 not only because of its relatively smaller loading but also because it pertained to habitual reading of science news, whereas the other retained items were relevant to general tendencies to seek out evidence as they encounter new information or claims. Among the five remaining items (Items 32, 34, 35, 41, 27), we employed (c) reliability criterion, which suggested the deletion of Item 35 would lower Cronbach's alpha the least. The deletion of Item 35 was conceptually reasonable as well, because it captured individuals' interest in "funding" source of information, whereas other retained items involved tendencies to examine or seek further evidence to check the validity of information or claims at hand. Following these decisions, Factor 2 consisted of Items 34, 32, 41, and 27.

For Factor 3 (Pro-authority 1), there were six items loaded onto the factor, where Items 3 and 14 conceptually overlapped – both items were about "parents" as authority figures. Between the two, we dropped Item 3 because its factor loading had smaller magnitude than that of Item 14. Among the remaining five items, to decide which one of the two items should be removed, we employed (c) reliability criterion, which suggested that dropping Item 24 would lower internal reliability of the scale the least. Among the items that loaded on Factor 3, Items 17 and 14 were originally intended to capture the Anti-authority reasoning tendency, thus their factor loadings suggested they be reverse-coded on this factor. It is possible in the exploratory stage of scale development that researchers may encounter some items that load onto a factor different from their original expectations, which requires considerations of whether the set of items that load together share a conceptual meaning that reasonably constitutes a single construct (e.g.,

Newton et al., 2021; Svedholm-Häkkinen & Lindeman, 2017). We reasoned that this factor would make a conceptual sense as a scale that captures individuals' tendency to rely on and follow authority figures if Items 17 and 14 are reverse-coded. However, relatively complex nature of Factor 3 led us to consider Factor 5, on which the items we originally designed to capture Pro-authority tendency loaded together, as a better candidate for Pro-authority dimension of the EvA scale. Thus, we tentatively named Factor 3 as Pro-authority 1, to compare with Factor 5, which we named as Pro-authority 2.

For Factor 4 (Anti-authority), five items were loaded on this factor. The (c) reliability criterion suggested that dropping Item 16, which had the smallest factor loading, would lower internal reliability of the scale the least. The remaining four items conceptually had a shared meaning – a tendency to resist relying on others' opinions. One of the items, Item 29, was originally designed to capture Pro-evidence tendency, but our reassessment of the item in light of other items on this factor suggested that this item was closely related to Anti-authority tendency, tapping onto the tendency to resist opinion-based news reports. Based on these considerations, Items 29, 21, 19, 23 consisted Factor 4.

Factor 5 (Pro-authority 2) and Factor 6 (Anti-medicine, Anti-evidence 2) contained four items that meaningfully loaded on each factor. Four items that loaded on Factor 5 (Items 2, 6, 9, 4) were all relevant to individuals' tendency to rely on and follow authority figures, such as politician, friends, and celebrities. Four items on Factor 6 (Items 44, 45, 43, 54), although they were originally developed to capture Anti-evidence tendency, were specifically related to individuals' aversion to medicine and chemical.

Table S8

Exploratory Factor Analysis: Factor Loadings of Four Items Selected per Factor (Subset of Table S6)

Item	Statement	Expected Trait	Factors					
			1	2	3	4	5	6
57	**Scientific evidence is overrated; there are often better ways to understand the world. (E-)**	**Anti-evidence 1**	.84					
50	**Even if scientific studies are done carefully and transparently, I still don't really believe them. (E-)**		.76					
48	**People make too much of scientific studies in the news when I know that the research is biased anyway. (E-)**		.62					
47	**When new evidence reverses a previous scientific theory, I just stop paying attention to it and make my own decisions. (E-)**		.57					
32	**When I hear a news story reporting research about health, I want to look up the study they are referring to. (E+)**	**Pro-evidence**		.85				
34	**I carefully examine research on important issues to make sure it is valid and unbiased. (E+)**			.84				
41	**When debating an important issue, I try to fact-check things that people state as statistics. (E+)**			.68				
27	**When someone makes a statement that sounds like a fact, I want to know the evidence behind it. (E+)**			.67				
5	I respect law enforcement, like police officers. (A)	Pro-authority 1			.77			
12	People should always respect authority. (A, adapted from UDS)				.69			
14	Children should be allowed to question their parents' authority. (A-, adapted from EBI)				-.56			
17	I'm the type of person who questions authority. (A-, adapted from UDS)				-.61			

29	**I am concerned that news reports are based on people's opinions rather than actual evidence. (A-)**	**Anti-authority**				.66		
21	**Government officials often say things that are untrue in their public statements. (A-)**					.66		
19	**Hosts of major television news shows do not know enough to be reliable sources of information. (A-)**					.61		
23	**People who are telling us how to act don't always have an incentive to tell the truth. (A-)**					.49		
2	**When I have to vote, I see what my politician says and follow their lead. (A)**	**Pro-authority 2**					.64	
6	**I often make changes to my diet based on what my friends tell me is more healthy. (A)**						.55	
9	**When I think a politician has a confident, assertive personality, I naturally like them and vote for them. (A)**						.54	
4	**I assume that when my favorite blogger or social media personality gives advice, they know what they are talking about. (A)**						.51	
44	I am wary of medical procedures that interfere with my body's natural processes. (E-)	Anti-evidence 2						.79
45	I am hesitant to take prescription medicines because they seem like chemicals I am putting into my body. (E-)							.76
43	I believe that things that are natural are always better for you. (E-)							.63
54	I avoid GMOs and pesticides, no matter what the evidence says. (E-)							.53
		Alpha	.85	.82	.78	.74	.68	.76

Note. Entries are standardized factor loadings. Factor loadings smaller than .3 are not displayed. Bolded items were retained; Letters in parentheses indicate *a priori* dimensions from item development: E+ (Pro-evidence), E- (Anti-evidence), A+ (Pro-authority), A- (Anti-authority).

Each subscale had acceptable internal reliability (Cronbach's alpha): Factor 1 (Anti-evidence): .85; Factor 2 (Pro-evidence): .82; Factor 3 (Pro-authority 1): .78; Factor 4 (Anti-authority): .74, Factor 5 (Pro-authority 2): .68; Factor 6 (Anti-medicine): .76 (DeVellis, 2017; Tavakol & Dennick, 2011). Based on the rationales for factor retention that we explain in the main text, we retained Factors 1, 2, 4 and 5. For validation in Study 2, we retained four items per factor chosen based on the considerations above, in order to create a smaller, more efficient 16-item scale that was balanced by factor.

In the main text, we also use AIC to assess the model fit. AIC adjusts χ^2 for the number of estimated parameters, allowing us to compare non-nested competing models, with lower AIC suggesting a better model fit (Schermelleh-Engel et al., 2003). The conceptual distinctness of our scale compared to other similar prior ones is bolstered by the fact that all of items taken from existing scales in our initial, larger set were not retained in the final scale.

Studies 1-3 Results

Factor Correlations

Table S9

Correlations among the EvA Subscales: Studies 1-3

	1	2	3	4
Study 1 (N = 547)				
1. Pro-evidence	(.82)			
2. Anti-evidence	–.19***	(.85)		
3. Pro-authority	–.12**	.36***	(.68)	
4. Anti-authority	.27***	.24***	–.22***	(.74)
Study 2 (N = 189)				
1. Pro-evidence	(.84)			
2. Anti-evidence	–.19***	(.83)		
3. Pro-authority	–.06	.32***	(.75)	
4. Anti-authority	.36***	.24***	–.28***	(.68)
Study 3 (N = 316)				
1. Pro-evidence	(.75)			
2. Anti-evidence	–.08	(.84)		
3. Pro-authority	.04	.52***	(.78)	
4. Anti-authority	.26***	.12**	–.20***	(.68)

Note. Entries are bivariate correlations among EvA subscales with coefficient alphas on the diagonal in parentheses. ***p < .01, **p < .05, *p < .10.

Confirmatory Factor Analysis

Table S10

Confirmatory Factor Analysis and Item-total Correlations: Studies 1-3

	Study 1		Study 2		Study 3	
	Factor loadings	Item-total correlation	Factor loadings	Item-total correlation	Factor loadings	Item-total correlation
Pro-evidence						
When I hear a news story reporting research about health, I want to look up the study they are referring to.	.70	.63	.67	.74	.65	.54
I carefully examine research on important issues to make sure it is valid and unbiased.	.76	.68	.78	.66	.66	.56
When debating an important issue, I try to fact-check things that people state as statistics	.72	.61	.75	.64	.70	.59
When someone makes a statement that sounds like a fact, I want to know the evidence behind it.	.78	.67	.83	.57	.65	.53
Anti-evidence						
Scientific evidence is overrated; there are often better ways to understand the world.	.77	.70	.85	.61	.83	.73
Even if scientific studies are done carefully and transparently, I still don't really believe them.	.81	.73	.76	.72	.73	.66
People make too much of scientific studies in the news when I know that the research is biased anyway.	.78	.68	.70	.64	.75	.67

When new evidence reverses a previous scientific theory, I just stop paying attention to it and make my own decisions.	.72	.64	.63	.71	.70	.61
Pro-authority						
When I have to vote, I see what my politician says and follow their lead.	.65	.53	.68	.44	.70	.60
I often make changes to my diet based on what my friends tell me is more healthy.	.53	.42	.61	.52	.60	.53
When I think a politician has a confident, assertive personality, I naturally like them and vote for them.	.60	.45	.69	.50	.81	.65
I assume that when my favorite blogger or social media personality gives advice, they know what they are talking about.	.59	.44	.64	.40	.61	.54
Anti-authority						
I am concerned that news reports are based on people's opinions rather than actual evidence.	.72	.59	.55	.57	.59	.48
Government officials often say things that are untrue in their public statements.	.66	.57	.67	.51	.61	.47
Hosts of major television news shows do not know enough to be reliable sources of information.	.69	.55	.64	.57	.56	.46
People who are telling us how to act don't always have an incentive to tell the truth.	.50	.42	.52	.52	.60	.46
CFA fit statistics						
CFI	.954		.928		0.928	
TLI	.943		.911		0.912	
SRMR	.046		.069		0.071	
RMSEA	.049		.063		0.060	
$\chi^2(df)$	228.89 (98)		172.35 (98)		209.99 (98)	
N	547		189		316	

Note. Entries for factor loadings are standardized and all were statistically significant ($p < .01$).

For scale homogeneity, items are retained with item-total correlations above .3 (Streiner et al., 2015; e.g., Duckworth et al., 2007; Lipkus et al., 2001) and dropped for being below .2 (Ames et al., 2005; Morof et al., 2012). Across studies, all items contributed to the homogeneity of relevant underlying constructs.

Model Comparisons: Fit Statistics

Table S11

Fit Statistics for the Proposed and Alternative Models: Studies 1-3

	$\chi^2(df)$	$\chi^2_{diff}(\Delta df)$	RMSEA	SRMR	CFI	TLI
Study 1 (N = 547)						
Proposed model	228.89 (98)		.049	.046	.954	.943
Alternative model A	477.37 (104)	248.48 (6)***	.081	.132	.867	.847
Alternative model B	1725.64 (104)	1496.8 (6)***	.169	.174	.424	.336
Alternative model C	930.72 (103)	701.8 (5)***	.121	.123	.706	.658
Alternative model D	367.10 (103)	138.21 (5)***	.068	.095	.906	.891
Study 2 (N = 189)						
Proposed model	172.35 (98)		.063	.069	.928	.911
Alternative model A	293.67 (104)	121.32 (6)***	.098	.149	.815	.787
Alternative model B	763.39 (104)	591.04 (6)***	.183	.194	.358	.259
Alternative model C	428.89 (103)	256.54 (5)***	.129	.134	.683	.630
Alternative model D	243.85 (103)	71.5 (5)***	.085	.106	.863	.840

Study 3 (N = 316)						
Proposed model	209.99 (98)		.060	.071	.928	.912
Alternative model A	384.33 (104)	174.34 (6)***	.092	.150	.819	.791
Alternative model B	868.69 (104)	658.7 (6)***	.153	.152	.507	.431
Alternative model C	538.17 (103)	328.18 (5)***	.135	.144	.617	.553
Alternative model D	272.41 (103)	62.432 (5)***	.072	.092	.891	.873

Note. *Proposed model:* Items load on four factors (Pro-evidence, Anti-evidence, Pro-authority, Anti-authority); *A:* No relationships between factors; *B:* All items load on one factor; *C:* Items load on two factors (Evidence-oriented, Authority-oriented); *D:* Items load on four first-order factors, with two second-order factors (Evidence-oriented, Authority-oriented). ***p* < .1; ****p* < .05; *****p* < .01.

Study 3 Results

Correlations among the EvA subscales and Convergent/Discriminant Scales

Construct validity refers to the extent that an operationalization measures the construct it purports to measure (Campbell & Fiske, 1959). It can be assessed by whether the given measure is associated with other indicators in a way that conforms to theoretical expectations, through convergent and divergent validity. Convergent validity is established through a strong association with measures that are theoretically similar or overlapping whereas discriminant validity is achieved when theoretically distinct constructs are less associated (Adcock & Collier, 2001).

Table S12

Correlations among the EvA Subscales and Convergent/Discriminant Scales

		EvA scale			
		Pro-evidence	Anti-evidence	Pro-authority	Anti-authority
Convergent	Need for Cognition	.39***	–.11*	–.03	.06
	Distrust in Science	–.13**	.78***	.36***	.08
	Pro-Authoritarianism	.06	.63***	.64***	–.13**
	Anti-Authoritarianism	.03	–.28***	–.29***	.29***
Divergent	Numeracy	.05	–.31***	–.32***	.10*
	Pessimism	–.02	–.02	–.08	.22***
	Dispositional Trust	–.02	–.09	.04	–.23***
	Dispositional Distrust	.02	.09	–.04	.23***

Note. **p* < .1; ***p* < .05; ****p* < .01. The cells that evaluate convergent and discriminant validity of each EvA subscale are shaded.

Measurement Invariance and Internal Reliability of EvA Subscales among Subgroups

Table S13

Confirmatory factor analysis using the entire sample (N= 316), non-college sample (n=160), college sample (n=156)

	Factor											
	Pro-evidence			Anti-evidence			Pro-authority			Anti-authority		
Item	All	No college	College	All	No college	College	All	No college	College	All	No college	College
1	.65	.67	.59									
2	.66	.70	.61									
3	.70	.69	.70									
4	.65	.67	.67									
5				.83	.79	.82						
6				.73	.81	.68						
7				.75	.67	.76						
8				.70	.62	.70						
9							.70	.61	.72			
10							.60	.43	.64			
11							.81	.78	.81			
12							.61	.42	.68			
13										.59	.59	.60
14										.61	.59	.59
15										.56	.47	.67
16										.60	.61	.60
Alpha	.75	.77	.73	.84	.81	.83	.78	.65	.80	.68	.65	.71
M	.75	.73	.77	.33	.26	.41	.43	.35	.50	.68	.69	.67
SD	.15	.16	.14	.23	.19	.24	.22	.18	.23	.17	.16	.18

Note. All of the standardized factor loadings are statistically significant ($p < 0.01$). *M* = mean; *SD* = standard deviation. Means and standard deviations are based on coarse factor scores (i.e., composite score as the average of four items, scaled to range from 0 to 1).

Table S14

Tests for Measurement Invariance between Education Groups

Model	χ^2 (df)	χ^2_{diff} (Δdf)	RMSEA	SRMR	CFI	TLI
Configural invariance	322.56 (196)		0.064	0.071	0.913	0.893
Metric invariance	335.48 (208)	12.92 (12)	0.062	0.075	0.912	0.898
Scalar invariance	352.06 (220)	16.58 (12)	0.062	0.076	0.909	0.900
Residual invariance	389.74 (236)	37.68 (16)***	0.064	0.080	0.894	0.892

Note. *$p < .1$; **$p < .05$; ***$p < .01$; non-college = 160, college = 156.

Table S15

Confirmatory factor analysis using the entire sample (N= 316), female sample (n=145), male sample (n=162).

	Factor											
	Pro-evidence			Anti-evidence			Pro-authority			Anti-authority		
Item	All	Female	Male	All	Female	Male	All	Female	Male	All	Female	Male

1	.65	.67	.64									
2	.66	.62	.71									
3	.70	.70	.69									
4	.65	.63	.70									
5				.83	.86	.83						
6				.73	.66	.76						
7				.75	.65	.82						
8				.70	.75	.67						
9							.70	.72	.69			
10							.60	.55	.63			
11							.81	.76	.82			
12							.61	.60	.60			
13										.59	.57	.63
14										.61	.59	.62
15										.56	.54	.61
16										.60	.68	.52
Alpha	.75	.74	.77	.84	.82	.85	.78	.76	.78	.68	.69	.68
M	.75	.74	.76	0.33	0.31	0.36	0.43	0.38	0.47	0.68	0.68	0.69
SD	.15	.15	.14	0.23	0.21	0.24	0.22	0.20	0.22	0.17	0.17	0.17

Note. All of the standardized factor loadings are statistically significant ($p < 0.01$). *M* = mean; *SD* = standard deviation. Means and standard deviations are based on coarse factor scores (i.e., composite score as the average of four items, scaled to range from 0 to 1).

Table S16

Tests for Measurement Invariance between Gender Groups

Model	χ^2 (df)	χ^2_{diff} (Δdf)	RMSEA	SRMR	CFI	TLI
Configural invariance	321.26 (196)		0.065	0.077	0.918	0.900
Metric invariance	336.51 (208)	15.24 (12)	0.063	0.081	0.916	0.903
Scalar invariance	347.53 (220)	11.02 (12)	0.061	0.083	0.917	0.909
Residual invariance	371.67 (236)	24.14 (16)*	0.061	0.085	0.912	0.910

Note. *$p < .1$; **$p < .05$; ***$p < .01$; female = 145, male = 162.

In addition to the invariance tests in the main text of our paper, we additionally examined the final step of *residual (strict) measurement invariance*, which examines whether the sum of specific variances (variance of the items not shared with the factor) and error variance (measurement error) are similar between groups. For education groups, the equality constraints on item residuals significantly degraded the fit and so did not support residual invariance, $\chi^2_{diff}(16) = 37.68$, $p < 0.01$. Residual invariance was not supported between gender groups as well, because the equality constraints item residuals somewhat degraded model fit, $\chi^2_{diff}(16) = 24.14$, $p = 0.09$. However, because residuals are not part of the latent factor and because residual invariance could be too strict and unrealistic for group comparisons, the residual invariance is not a prerequisite for latent mean comparisons (Steinmetz et al., 2009; Putnick & Borstein, 2016).

Cronbach's α of some of the EvA subscales for certain subgroups is slightly lower than the recommended value of .7. However, this can be evaluated as acceptable given that the scale consists of the relatively low number of items (4 items per scale) and that the 'unacceptable level' is defined as Cronbach's alpha 'below .60.' (Peterson, 1994; Price & Mueller, 1986).

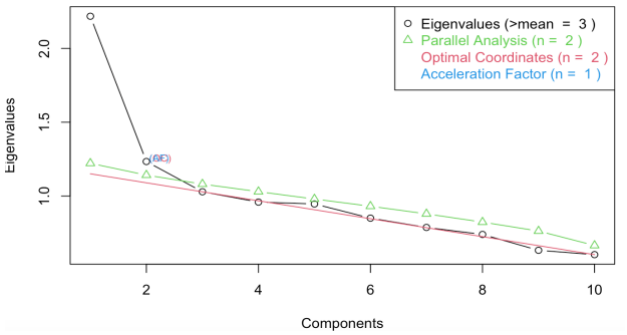
Cronbach (1951) suggested that a high value of alpha is 'desirable', but also made the point that the key point should be that scores obtained when using an instrument had to be interpretable—and this was often possible without needing very high values of alpha. Although adding more items into an instrument can increase the value of alpha, he pointed out that adding additional items that measure the same thing may result in a redundancy that is inefficient (Taber 2018, p.1288). Schmitt (1996) also suggested that "there is no general level where alpha becomes acceptable, but rather that instruments with quite a low value of alpha can still prove useful in some circumstances."

Measurement Issue with Social Desirability

In Study 3, we used a 10-item scale that was developed by Strahan & Gerbasi (1972), as suggested by Fischer & Fick (1993) (items are displayed in Table S17). Only after we conducted Study 3, we noticed that Ramanaiah & Martin (1980) suggested two components of social desirability: 1) Attribution: The tendency to attribute socially desirable characteristics; 2) Denial: The tendency to deny socially undesirable characteristics. Given the controversy on the dimensionality of social desirability in the literature (e.g., Hart et al., 2015; Helmes & Holden, 2003), we used a parallel analysis to assess how the items are related to the underlying construct, which suggested that there exists two underlying factors.

Figure S2

Parallel Analysis of the Social Desirability Items



We then used exploratory factor analysis with an oblique rotation to assess how the items loaded on two different factors. It suggested two sets of items that loaded on Attribution (Factor 1) and Denial (Factor 2) respectively, similar to what Ramanaiah & Martin (1980) suggested. Although this two-factor solution could be due to the method effect (i.e., the reverse-coded nature of half of the items), it was still tricky to figure out how we should construct the scale, particularly because one of the items on Factor 1 had relatively weak loading (< .4), and three of the items did not meaningfully load on either factor. While Fischer & Fick (1993) recommended the Strahan & Gerbasi (1972)'s scale, other studies have argued that Strahan & Gerbasi's scale might be less reliable than other measures of social desirability (e.g., Reynolds, 1982). Some have suggested that older measures of social desriability may not suit contemporary society, and have thus suggested using alternative set of items (Stöber, 2001, Hart et al., 2015).

Table S17

EFA on social desirability items in Study 3

Item	Statement	Factor 1	Factor 2
1	I'm always willing to admit it when I make a mistake.		0.42
2	I always try to practice what I preach.		
3	I never resent being asked to return a favor.		
4	I have never been irked when people expressed ideas very different from my own.		0.69
5	I have never deliberately said something that hurt someone's feelings.		0.42
6	I like to gossip at times. (R)		
7	There have been occasions when I took advantage of someone. (R)	0.64	
8	I sometimes try to get even rather than forgive and forget. (R)	0.36	
9	At times I have really insisted on having things my own way. (R)	0.43	
10	There have been occasions when I felt like smashing things. (R)	0.51	

Note: Factor loadings smaller than .3 are not displayed.

Furthermore, we also noticed that the social desirability scale we used in Study 3 contained a number of argument-related items (Items 1, 4, and 9), which could misleadingly inflate its correlation with some of our EvA scale. Therefore, we decided to administer an alternative measure of social desirability—based on a more recent study that identifies the two-factor structure of social desirability, in the context outside the realm of reasoning and argumentation (Impression Management dimension in Hart et al. 2015)— in Study 4 to more reliably test discriminant validity that the EvA traits relative to social desirability.

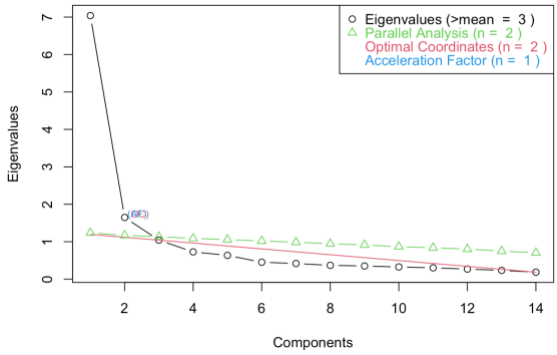
Study 4 Results

Exploratory Factor Analysis on Adherence to CDC Guidelines on COVID-19

In Study 4, we used fourteen items on CDC guideline adherence on COVID-19 by Graupensperger et al. (2021). Parallel analysis suggested that there exist two underlying factors. As shown in Table S18, the first factor is about distancing behaviors, whereas the second factor pertains to sanitizing behaviors. The necessity of sanitizing behaviors was less uniform across individuals than distancing behaviors. For instance, if individuals abide by CDC guidelines, they were expected to regularly engage with distancing behaviors. However, less frequency of sanitizing behaviors did not necessarily mean less compliance. For instance, individuals who mostly stay indoors had less need to engage with sanitizing behaviors, which did not necessarily indicate less compliance to CDC guidelines. Thus, we used the composite score of ten distancing behavior items as the main measure of adherence to CDC guidelines on COVID-19.

Figure S3

Parallel Analysis of the CDC Guideline Adherence on COVID-19



Exploratory factor analysis revealed that these fourteen items loaded onto two factors, distancing behaviors (ten items: e.g., six-feet distancing, wearing mask) and sanitizing behaviors (three items: e.g., hand washing, disinfecting surfaces); only one item, "getting tested when feeling sick," did not meaningfully load on either (Table S18).

Table S18

EFA on CDC Guideline Adherence on COVID-19 Items

Item	Statement	Factor 1	Factor 2
1	Hand washing with soap and water for 20 seconds when available		0.65
2	Using hand sanitizer in between activities		0.87
3	Wearing a face mask when in indoor public spaces (e.g., shopping)	0.58	
4	Staying 6 feet away from other people you don't live with	0.67	
5	Doing things at home rather than in public, when possible (e.g., work)	0.79	
6	Clean and disinfect frequently touched surfaces (e.g., tables, doorknobs)		0.84
7	Avoiding dining in restaurants by cooking meals at home and using takeout/delivery options	0.85	
8	Avoiding crowded indoor hang-out spots (e.g., bars, pubs, lounges)	0.96	
9	Avoiding large indoor gatherings such as weddings, shows, or parties	0.98	
10	Avoiding indoor social gatherings (e.g., friends' houses)	0.88	
11	Staying home and getting tested when feeling sick		
12	Avoiding contact with at-risk individuals (e.g., older people)	0.52	
13	Avoiding physical contact with others you do not live with (e.g., handshakes, hugs)	0.71	
14	Wearing a face mask while using public transportation (e.g., buses, trains, planes)	0.41	

Note: Factor loadings smaller than .3 are not displayed.

Robustness Check for the Criterion Validity Test

Following the preregistration, we additionally run OLS regression models on the relationships between the EvA scale and criterion variables (susceptibility to misinformation, adherence to social distancing, confidence in COVID vaccine, science curiosity, and religiosity) while controlling for demographic variables. The direction and statistical significance of the relationships largely remained the same even after controlling for age, gender, education, and income.

Table S19

Relationships between the EvA tendencies and Criterion Behaviors/Attitudes (Robustness check with control variables)

EvA Reasoning Tendencies	Perceived accuracy of health misinformation		Adherence to the CDC guide on social distancing		Confidence in COVID vaccine	
	b (SE)	*t*	*b (SE)*	*t*	*b (SE)*	*t*
Pro-evidence	0.02 (.05)	0.4	0.25 (0.08)	3.0***	0.19 (0.07)	2.5**
Anti-evidence	0.40 (.04)	9.8***	–0.25 (0.06)	–4.1***	–0.78 (0.06)	–14.0***
Pro-authority	0.16 (.05)	3.5***	0.03 (0.07)	0.4	0.37 (0.0)	5.4***
Anti-authority	–0.06 (.05)	–1.2	–0.12 (0.07)	–1.8*	–0.20 (0.07)	–2.5***
Age	–0.00 (.001)	–0.1	0.003 (0.001)	3.6***	–0.0002 (0.001)	–0.2
Female	0.02 (.01)	1.1	0.01 (0.02)	0.6	0.01 (0.02)	0.6
Education	–0.004 (.01)	–0.8	0.004 (0.01)	0.4	0.01 (0.01)	1.7*
Income	0.002 (0.002)	1.1	–0.01 (0.003)	–1.6*	0.01 (0.003)	2.3**
Constant	0.09 (0.07)	1.3	0.60 (0.10)	6.1***	0.68 (0.10)	6.6***
N	524		524		524	
Adjusted R^2	.27		.10		.34	

EvA Reasoning Tendencies	Science Curiosity		Religiosity	
	b (SE)	*t*	*b (SE)*	*t*
Pro-evidence	0.63 (0.05)	11.8***	0.08 (0.08)	1.0
Anti-evidence	–0.14 (0.05)	–3.0***	0.41 (0.07)	5.9***
Pro-authority	0.08 (0.06)	1.4	0.24 (0.08)	2.9***
Anti-authority	–0.01 (0.06)	-0.1	–0.15 (0.08)	–1.9*
Age	0.001 (0.001)	0.9	0.002 (0.001)	2.8***
Female	–0.05 (0.02)	–2.8***	0.04 (0.02)	1.7*
Education	0.02 (0.01)	2.9***	0.01 (0.01)	0.9
Income	0.01 (0.003)	2.2***	0.01 (0.004)	1.6
Constant	–0.05 (0.07)	–0.7	-0.03 (0.10)	-0.3
N	524		524	
Adjusted R^2	.26		.15	

Note. *b* = unstandardized regression coefficient robust with standard errors in parentheses. *t* = t-value for regression coefficient. To facilitate comparisons of coefficients, the four EvA tendencies and dependent variables were scaled to range from 0 to 1. *Age* ranged between 18 and 78. *Female* was a binary variable, 1 if female, 0 if male. *Education* and *Income* were coded as shown in Table S5. *$p < .10$; **$p < .05$; ***$p < .01$.

Relationships between the EvA Scale and Additional Criterion Variables

We examined the relationships between the EvA scale and additional variables available in our study. Some of these variables were preregistered as exploratory analyses (e.g., trust in various sources of COVID-related information). The questionnaires designed to measure susceptibility to misinformation (Scherer et al. 2021) included not only social media posts with false health information, but also those with true health information to prevent subjects from making inferences about the research purpose. While our main analysis focused on perceived accuracy of *false* health information (susceptibility to health misinformation), as a post-hoc analysis, we additionally examined the relationships between the EvA traits and other available behaviors regarding health information. The relationships between the EvA traits and COVID-related behaviors and attitudes (e.g., vaccine intake, trust in various sources of COVID-related

information) were preregistered as exploratory analysis. Overall, these additional results indicate the four EvA reasoning tendencies identify individual differences in attitudes and decision making on the basis of evidence versus authority.

Table S20

Relationships between the EvA tendencies and Criterion Behaviors/Attitudes: Perceived Accuracy and Decision Influence of Health Information, COVID vaccine intake

EvA Reasoning Tendencies	Perceived accuracy of *true* info	Decision influence of *true* info	Decision influence of *false* info	COVID vaccine intake[1]
Pro-evidence	0.11 (0.04)***	0.14 (0.06)**	0.006 (0.05)	–0.06 (0.27)
Anti-evidence	–0.29 (0.04)***	–0.18 (0.06)***	0.035 (0.04)***	–2.09 (0.21)***
Pro-authority	–0.02 (0.04)	0.23 (0.06)***	0.22 (0.05)***	0.18 (0.25)
Anti-authority	0.08 (0.05)	–0.07 (0.06)	–0.06 (0.05)	–0.11 (0.28)
Constant	0.72 (0.05)***	0.51 (0.06)***	0.04 (0.06)	2.68 (0.31)***
N	529	529	529	529
Adjusted R^2	0.17	0.06	0.24	0.16

Note. Unstandardized regression coefficient with robust standard errors in parentheses. All variables were scaled to range from 0 to 1. *$p < .10$; **$p < .05$; ***$p < .01$. *Perceived accuracy* of true health information was measured as the average of perceived accuracy ratings on the four true headlines. *Decision influence* refers to the degree to which individuals rated the given information would influence their own cancer treatment decisions. Decision influence variables were constructed as the average of four *true* and *four* false social media posts respectively. *COVID vaccine intake* refers to the number of COVID vaccine doses, ranging from 0 to 3.

Table S21

Relationships between the EvA tendencies and Criterion Behaviors/Attitudes: Trust in Sources of COVID-related Information

EvA Reasoning Tendencies	Trust in Centers for Disease Control and Prevention (CDC)	Trust in Food and Drug Administration (FDA)	Trust in religious leaders	Trust in social media
Pro-evidence	0.08 (0.07)	0.08 (0.07)	–0.05 (0.07)	0.004 (0.06)
Anti-evidence	–0.69 (0.06)***	–0.51 (0.06)***	0.32 (0.06)***	0.13 (0.05)***
Pro-authority	0.41 (0.07)***	0.42 (0.06)***	0.44 (0.07)***	0.55 (0.05)***
Anti-authority	–0.29 (0.07)***	–0.37 (0.08)***	–0.25 (0.07)***	–0.29 (0.06)***
Constant	0.89 (0.07)***	0.85 (0.07)***	0.23 (0.07)***	0.20 (0.07)***
N	529	529	529	529
Adjusted R^2	0.32	0.27	0.24	

Note. Unstandardized regression coefficient with robust standard errors in parentheses. All variables were scaled to range from 0 to 1. *$p < .10$; **$p < .05$; ***$p < .01$. *Trust in [CDC/FDA/religious leaders/social media]* refers to the degree of trust in [CDC/FDA/religious leaders/social media] as the source of information about COVID-19 vaccines on a five-point scale ranging from "strongly distrust" to "strongly trust."

[1] The direction and statistical significance of the relationships between the EvA traits and COVID vaccine intake remains the same when we run a logistic regression by using a binary variable that indicates COVID vaccine intake (0 = did not take any, 1 = took at least one COVID vaccine dose).

Survey Instruments

Note: The study materials, raw data, and R code for this study are available at: https://osf.io/qeav5/?view_only=cc20ea01c47b48adbde31f7e1e7ec52d.

Study 1

- **Initial 57 items for the scale development**

Instruction:

> We'd like to ask you questions about your typical preferences when obtaining or processing information. Please indicate how much you disagree or agree with each statement.

Response options:

> Strongly disagree (1) – Disagree (2) – Somewhat disagree (3) – Neither disagree nor agree (4) – Somewhat agree (5) – Agree (6) – Strongly Agree (7)

Note: The order of items was randomized.

Item	Statement
1	I pay close attention to what my religious leader tells me I should do.
2	When I have to vote, I see what my politician says and follow their lead.
3	I generally follow my parents' advice.
4	I assume that when my favorite blogger or social media personality gives advice, they know what they are talking about.
5	I respect law enforcement, like police officers.
6	I often make changes to my diet based on what my friends tell me is more healthy.
7	My behavior is usually dictated by my religious values.
8	I assume my doctors know what they're talking about, so I follow their recommendations.
9	When I think a politician has a confident, assertive personality, I naturally like them and vote for them.
10	People who challenge authority are overconfident.
11	Sometimes you just have to accept the teacher's answer even though you don't understand it.
12	People should always respect authority.
13	People who are in a position of authority have the right to tell others what to do.
14	Children should be allowed to question their parents' authority.
15	I wonder how much my teachers really knew.
16	Even advice from experts should be questioned.
17	I'm the type of person who questions authority.
18	I am doubtful that my teachers really understood what they were teaching me.
19	Hosts of major television news shows do not know enough to be reliable sources of information.
20	Just because people are older or more experienced does not mean their claims are necessarily correct.
21	Government officials often say things that are untrue in their public statements.

22	I assume that people in positions of power are corrupt.
23	People who are telling us how to act don't always have an incentive to tell the truth.
24	From my perspective, people in positions of authority should generally not be trusted.
25	Scientists' research doesn't matter in the real world.
26	When I hear a news story about health, I wonder if there is really good evidence behind the assertion.
27	When someone makes a statement that sounds like a fact, I want to know the evidence behind it.
28	When someone cites a statistic, I want to know where they got it from.
29	I am concerned that news reports are based on people's opinions rather than actual evidence.
30	I pay attention to science news and try to follow the latest findings.
31	I am more likely to avoid a risk when I learn about the statistics rather than personal stories and anecdotes.
32	When I hear a news story reporting research about health, I want to look up the study they are referring to.
33	Before I vote on an issue in my state or city, I try to look up the ballot items so that I vote correctly.
34	I carefully examine research on important issues to make sure it is valid and unbiased.
35	When I hear about new research, I look into who funded it to be sure it is unbiased.
36	I think news reports about science should include more information so that we can evaluate the strength of the evidence.
37	When my doctor tells me about a new treatment, I like to find out about any research on the treatment.
38	Science is very important for the country's development.
39	Money spent on science is well worth spending.
40	In general, you should consider whether the information in your textbook is accurate.
41	When debating an important issue, I try to fact-check things that people state as statistics.
42	It is usually wise to seek out evidence and research before making decisions
43	I believe that things that are natural are always better for you.
44	I am wary of medical procedures that interfere with my body's natural processes.
45	I am hesitant to take prescription medicines because they seem like chemicals I am putting into my body.
46	When scientists change their minds, I stop trusting their research on what we are supposed to eat to be healthy.
47	When new evidence reverses a previous scientific theory, I just stop paying attention to it and make my own decisions.
48	People make too much of scientific studies in the news when I know that the research is biased anyway.
49	People can talk about data, but I think that my intuitions are a better guide for my decisions.
50	Even if scientific studies are done carefully and transparently, I still don't really believe them.
51	I am not interested in looking into the details when I hear the results of a new study.
52	I think scientific data is too hard to understand, so I generally ignore it.
53	When it comes to controversial issues in society, I don't think "the data" can tell us much.
54	I avoid GMOs and pesticides, no matter what the evidence says.
55	I rarely check the nutrition facts or ingredient list on food labels.
56	When choosing between products, I don't spend much time comparing the specifications.

57	Scientific evidence is overrated; there are often better ways to understand the world.

Study 2

- **The 16-item Evidence versus Authority (EvA) scale**

Instruction:

> We'd like to ask you questions about your typical preferences when obtaining or processing information. Please indicate how much you disagree or agree with each statement.

Response options:

> Strongly disagree (1) – Disagree (2) – Somewhat disagree (3) – Neither disagree nor agree (4) – Somewhat agree (5) – Agree (6) – Strongly Agree (7)

Note: The order of items was randomized.

Pro-evidence	
1	When I hear a news story reporting research about health, I want to look up the study they are referring to.
2	I carefully examine research on important issues to make sure it is valid and unbiased.
3	When debating an important issue, I try to fact-check things that people state as statistics
4	When someone makes a statement that sounds like a fact, I want to know the evidence behind it.
Anti-evidence	
5	Scientific evidence is overrated; there are often better ways to understand the world.
6	Even if scientific studies are done carefully and transparently, I still don't really believe them.
7	People make too much of scientific studies in the news when I know that the research is biased anyway.
8	When new evidence reverses a previous scientific theory, I just stop paying attention to it and make my own decisions.
Pro-authority	
9	When I have to vote, I see what my politician says and follow their lead.
10	I often make changes to my diet based on what my friends tell me is more healthy.
11	When I think a politician has a confident, assertive personality, I naturally like them and vote for them.
12	I assume that when my favorite blogger or social media personality gives advice, they know what they are talking about.
Anti-authority	
13	I am concerned that news reports are based on people's opinions rather than actual evidence.
14	Government officials often say things that are untrue in their public statements.
15	Hosts of major television news shows do not know enough to be reliable sources of information.
16	People who are telling us how to act don't always have an incentive to tell the truth.

Study 3

Note: For each scale, the order of items was randomized.

- **Need for Cognition Scale** (NCS; Coelho et al., 2018)

Response options:

Strongly disagree (1) – Disagree (2) – Somewhat disagree (3) – Neither disagree nor agree (4) – Somewhat agree (5) – Agree (6) – Strongly Agree (7)

Statements:

- I would prefer complex to simple problems.
- I like to have the responsibility of handling a situation that requires a lot of thinking.
- Thinking is not my idea of fun. (R)
- I would rather do something that requires little thought than something that is sure to challenge my thinking abilities. (R)
- I really enjoy a task that involves coming up with new solutions to problems.
- I would prefer a task that is intellectual, difficult, and important to one that is somewhat important but does not require much thought.

- **Distrust in Science** (Nadelson et al., 2020; 12 distrust items)

Response options:

Strongly disagree (1) – Disagree (2) – Neutral (3) – Agree (4) – Strongly Agree (5)

Statements:

- When scientists change their mind about a scientific idea it diminishes my trust in their work.
- Scientists ignore evidence that contradicts their work.
- Scientific theories are weak explanations.
- Scientists intentionally keep their work secret.
- Scientists don't value the ideas of others.
- Scientists don't care if laypersons understand their work.
- When scientists form a hypothesis they are just guessing.
- We cannot trust scientists because they are biased in their perspectives.
- Scientist will protect each other even when they are wrong.
- We cannot trust scientists to consider ideas that contradict their own.
- Today's scientists will sacrifice the well being of others to advance their research.
- We cannot trust science because it moves too slowly.

- **Right-Wing Authoritarianism** (RWA; Bizumic & Duckitt, 2018)

Response options:

Strongly disagree (1) – Disagree (2) – Somewhat disagree (3) – Neither disagree nor agree (4) – Somewhat agree (5) – Agree (6) – Strongly Agree (7)

- **Pro-Authoritarianism**
  - What our country needs most is discipline, with everyone following our leaders in unity.
  - God's laws about abortion, pornography, and marriage must be strictly followed before it is too late.

  - The facts on crime and the recent public disorders show we have to crack down harder on troublemakers, if we are going to preserve law and order.

- **Anti-Authoritarianism**
  - It's great that many young people today are prepared to defy authority.
  - There is nothing wrong with premarital sexual intercourse.
  - Our society does NOT need tougher government and stricter laws.

- **Numeracy** (Weller et al., 2013)

Instruction:

Please answer the following questions by entering your answer into the box using numbers only.

Questions:

- Imagine that we roll a fair, six-sided die 1000 times. Out of 1000 rolls, how many times do you think the die would come up as an even number? ____
- In the BIG BUCKS LOTTERY, the chances of winning a $10.00 prize are 1%. What is your best guess about how many people would win a $10.00 prize if 1000 people each buy a single ticket from BIG BUCKS? ____
- In the ACME PUBLISHING SWEEPSTAKES, the chance of winning a car is 1 in 1000. What percent of tickets of ACME PUBLISHING SWEEPSTAKES win a car? ____ %
- If the chance of getting a disease is 10%, how many people would be expected to get the disease out of 1000 people? ____
- If the chance of getting a disease is 20 out of 100, this would be the same as having a ____% chance of getting the disease.
- A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost? ____ cents
- If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? ____ minutes
- In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake? ____ days

- **Pessimism** (Scheier et al., 2012)

Response options:

Disagree a lot (1) – Disagree a little (2) – Neither agree nor disagree (3) – Agree a little (4) – Agree a lot (5)

Statements:

- If something can go wrong for me, it will.
- I hardly ever expect things to go my way.
- I rarely count on good things happening to me.

- **Dispositional Trust/Distrust** (Bianchi & Brockner, 2012)

Do you think most people would try to take advantage of you if they got a chance or would they try to be fair?

Take advantage (1) - Depends (2) - Fair (3)

Generally speaking, would you say that most people can be trusted or that you can't be too careful in dealing with people?

Cannot trust (1) - Depends (2) - Can trust (3)

Would you say that most of the time people try to be helpful, or that they are mostly just looking out for themselves?

Look out for themselves (1) - Depends (2) - Helpful (3)

Study 4

- **Susceptibility to Health Misinformation** (Scherer et al., 2021)

Instruction:

> [Screen 1]
> Now, you will view pictures and information that have been shared publicly on social media.
> Your job is to evaluate the accuracy of each one. We are interested in your personal opinion.
> Some social media posts may contain multiple claims. In these cases, tell us what you think of it overall.
> Later, we will ask you some questions about yourself, and then you will be done.
> Click the arrow button for a few more instructions.
>
> [Screen 2]
> You will be judging social media posts about a health-related topics: cancer treatments
>
> [Screen 3]
> Now you are ready to begin. The first thing you see will be a social media post. Please read it and then answer the two questions below it.
> You will rate a total of 8 social media posts.
> Click the arrow button to start rating the social media posts.

Questionnaires for each social media post:

- To the best of your knowledge, how accurate is the information in this social media post?
  - Completely false (1) - Mostly false (2) - Mostly true (3) - Completely true (4)

- If you were diagnosed with cancer, would this information influence your decision about your treatment?
  - No, definitely not (1) - Probably not (2) - Probably yes (3) - Yes, definitely (4)

Social media posts:[2]

- Social media posts with FALSE information:

1

2

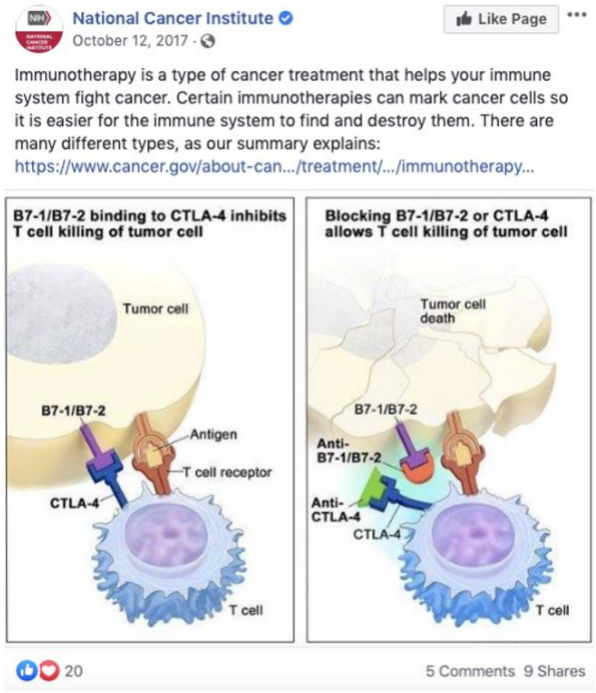
3

4

- Social media posts with TRUE information:

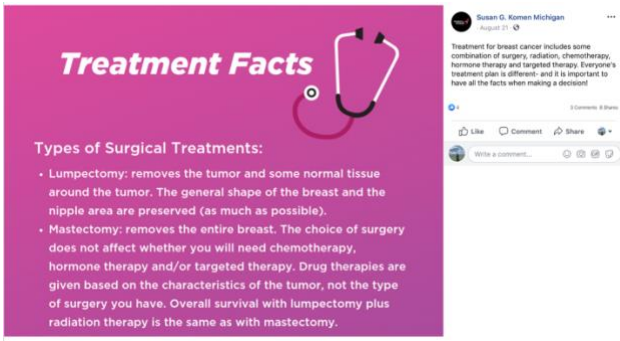
[2] These experimental stimuli were adopted from supplemental materials (cancer treatments) for Scherer et al. (2021).

1



2

3



4

- **Adherence to CDC Guidelines on COVID-19** (Graupensperger et al., 2021)

Instruction:

Please rate the degree to which you engaged in each activity during the past month.

Response options:

Never (1) – Rarely (2) – Sometimes (3) – Often (4) – All the time (5)

Statements:

- Hand washing with soap and water for 20 seconds when available
- Using hand sanitizer in between activities

- **Wearing a face mask when in indoor public spaces (e.g., shopping)**
- **Staying 6 feet away from other people you don't live with**
- **Doing things at home rather than in public, when possible (e.g., work)**
- Clean and disinfect frequently touched surfaces (e.g., tables, doorknobs)
- **Avoiding dining in restaurants by cooking meals at home and using takeout/delivery options**
- **Avoiding crowded indoor hang-out spots (e.g., bars, pubs, lounges)**
- **Avoiding large indoor gatherings such as weddings, shows, or parties**
- **Avoiding indoor social gatherings (e.g., friends' houses)**
- Staying home and getting tested when feeling sick
- **Avoiding contact with at-risk individuals (e.g., older people)**
- **Avoiding physical contact with others you do not live with (e.g., handshakes, hugs)**
- **Wearing a face mask while using public transportation (e.g., buses, trains, planes)**

Note: Bolded items were used to construct a measure of distancing behaviors (explained in Table S13).

- **COVID-19 Related Attitudes** (CDC, 2021)

Confidence in COVID-19 vaccine:

- How likely are you to recommend getting the COVID-19 vaccine to others?
  - Not at all likely (1) - A little likely (2) - Somewhat likely (3) - Very likely (4) - Extremely likely (5)

- How safe do you think a COVID-19 vaccine is for you?
  - Not at all safe (1) - A little safe (2) - Moderately safe (3) - Very safe (4) - Extremely safe (5)

- How much do you distrust or trust the public health agencies that recommend you get a COVID-19 vaccine?
  - Strongly distrust (1) - Distrust (2) - Neither distrust nor trust (3) - Trust (4) - Strongly trust (5)

- How much confidence do you have that the research and development process have produced COVID-19 vaccines in the U.S. that are safe and effective?
  - None at all (1) - A little (2) - A moderate amount (3) - A lot (4) - A great deal (5)

Trust in sources of COVID-19 information:

Instruction:

How much do you distrust or trust the following as sources of information about COVID-19 vaccines?

Note: These response options and items were presented as a matrix.

Response options:

Strongly distrust (1) - Distrust (2) - Neither distrust nor trust (3) - Trust (4) - Strongly trust (5)

Items:

- Centers for Disease Control and Prevention (CDC)
- Food and Drug Administration (FDA)
- Religious leader(s)
- Social media (e.g., Facebook, Twitter, Instagram, WhatsApp, LinkedIn, or TikTok)

Intake of COVID-19 vaccine:

Q1 Have you received a COVID-19 vaccine?

Yes (1) - No (2)

[Display if Q1==1] Q2 Did you receive a vaccine product that requires only one dose or two doses?

One-dose product (e.g., Johnson & Johnson) (1)
Two-dose product (e.g., Pfizer, Moderna) (2)
Mix and match (e.g., J&J and Pfizer, Moderna and Pfizer, etc.) (3)

[Display if Q1==1] Q3 How many doses of COVID vaccines have you gotten so far?

One dose (1) - Two doses (2) - Three doses (3)

- **Science Curiosity** (Kahan et al. 2017; Motta et al., 2021)

Note: Following Motta et al. (2021) and our preregistration, science curiosity was measured as a composite score of bolded items.

Q1 There are a lot of issues in the news and it is hard to keep up with every area. We will list some topics that are covered in the media.

Please indicate **how closely you follow news relating to each topic** either in the newspaper, on television, on the radio, or on the Internet.

	Not at all (1)	A little but not closely (2)	Closely but not very closely (3)	Very closely (4)
Government or politics (1)				
Religion (2)				
Scientific research or discoveries (3)				
New technologies (4)				
Entertainment or celebrities (5)				

Q2 We'd also like to know whether you read books in your spare time. We will list some book topics. Please indicate **whether you have read a book on that topic <u>in the previous year</u>**.

	Did not read any books on the topic in the previous year (1)	Have read **1 book** on the topic in the previous year (2)	Have read **2 books** on the topic in the previous year (3)	Have read **more than 3 books** on the topic in the previous year (4)
Science fiction (1)				
Mystery novel (2)				
Government or politics (3)				
Religion (other than holy script text) (4)				
Scientific research or discoveries (5)				

Q3 We are also interested in knowing about the sorts of topics you discuss with family members, friends, or co-workers. We will list some conversation topics. Please indicate **how often you discuss these topics with either friends, family members, or co-workers.**

	Never (1)	Rarely (2)	More than rarely but not often (3)	Often (4)
Government or politics (1)				
Religion (2)				
Scientific research or discoveries (3)				
New technologies (4)				
Entertainment or celebrities (5)				

Q4 We will now list some topics that some people are interested in, and some people are not interested in. For each topic, please indicate **how interested you are in that topic.**

	Not at all interested (1)	Slightly interested (2)	Moderately interested (3)	Very interested (4)
Government or politics (1)				
Religion (2)				
Scientific research or discoveries (3)				
New technologies (4)				
Entertainment or celebrities (5)				

- **Religiosity** (Rohrboaugh & Jessor, 1975)
    - How often have you attended religious services during the past year?

  - Never (1)
  - A few times a year (2)
  - Once or twice a month (3)
  - Almost every week (4)
  - Every week (5)

- How often do you pray or practice religious meditation?
  - Never (1)
  - Rarely (2)
  - Occasionally (3)
  - Fairly often (4)
  - Very often (5)

- When you have a serious personal problem, how often do you take religious advice or teaching into consideration?
  - Never (1)
  - Rarely (2)
  - Occasionally (3)
  - Fairly often (4)
  - Very often (5)

- How much influence would you say that religion has on the way that you choose to act and the way that you choose to spend your time each day?
  - No influence (1)
  - A small influence (2)
  - Some influence (3)
  - A fair amount of influence (4)
  - A large influence (5)

- Which of the following statements comes closest to your belief about God? (pick one)
  - I don't believe in a personal God or in a higher power (1)
  - I don't know if there is a personal God or a higher power of some kind, and I don't know if I ever will. (2)
  - I don't know if there is a personal God, but I do believe in a higher power of some kind. (3)
  - Although I sometimes question God's existence, I do believe in God and believe God knows of me as a person. (4)
  - I am sure that God really exists and is active in my life. (5)

- **Social Desirability** (Hart et al., 2015; Impression Management)

Response options:

Strongly disagree (1) – Disagree (2) – Slightly disagree (3) – Neither disagree nor agree (4) – Slightly agree (5) – Agree (6) – Strongly Agree (7)

Statements:

- I sometimes tell lies if I have to. (R)
- I never cover up my mistakes.
- There have been occasions when I have taken advantage of someone. (R)
- I sometimes try to get even rather than forgive and forget. (R)
- I have said something bad about a friend behind his or her back. (R)
- When I hear people talking privately, I avoid listening.
- I never take things that don't belong to me.
- I don't gossip about other people's business.

References

Adcock, R., & Collier, D. (2001). Measurement validity: A shared standard for qualitative and quantitative research. *The American Political Science Review*, *95*(3), 529–546.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*(2), 81–105.

Clark, L. A., & Watson, D. (2016). *Constructing validity: Basic issues in objective scale development* (p. 203). American Psychological Association.

DeVellis, R. F. (2006). Classical Test Theory. *Medical Care*, *44*(11), S50–S59.

Gastañaduy, P. A., Budd, J., Fisher, N., Redd, S. B., Fletcher, J., Miller, J., McFadden, D. J., Rota, J., Rota, P. A., Hickman, C., Fowler, B., Tatham, L., Wallace, G. S., de Fijter, S., Parker Fiebelkorn, A., & DiOrio, M. (2016). A measles outbreak in an underimmunized Amish community in Ohio. *New England Journal of Medicine*, *375*(14), 1343–1354.

Gliem, J. A., & Gliem, R. R. (2003). *Calculating, interpreting, and reporting Cronbach's alpha reliability coefficient for likert-type scales*. Presented at the Midwest Research-to-Practice Conference in Adult, Continuing, and Community Education. https://hdl.handle.net/1805/344

Helmes, E., & Holden, R. R. (2003). The construct of social desirability: One or two dimensions? *Personality and Individual Differences*, *34*(6), 1015–1023.

Peterson, R. A. (1994). A meta-analysis of Cronbach's coefficient alpha. *Journal of Consumer Research*, *21*(2), 381–391.

Price, J. L., & Mueller, C. W. (1986). *Handbook of organizational measurement*. Pitman: Longman.

Ramanaiah, N. V., & Martin, H. J. (1980). On the Two-Dimensional Nature of the Marlowe-Crowne Social Desirability Scale. *Journal of Personality Assessment*, *44*(5), 507–514.

Raubenheimer, J. (2004). An item selection procedure to maximize scale reliability and validity. *SA Journal of Industrial Psychology*, *30*(4), 59–64.

Raykov, T. (2008). Alpha if item deleted: A note on loss of criterion validity in scale development if maximizing coefficient alpha. *British Journal of Mathematical and Statistical Psychology*, *61*(2), 275–285.

Reynolds, W. M. (1982). Development of reliable and valid short forms of the Marlowe-Crowne Social Desirability Scale. *Journal of Clinical Psychology*, *38*(1), 119–125.

Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, *8*(4), 350–353.

Taber, K. S. (2018). The use of Cronbach's Alpha when developing and reporting research instruments in science education. *Research in Science Education*, *48*(6), 1273–1296.