Social Sciences & Humanities Open 13 (2026) 102469

Contents lists available at ScienceDirect

Social Sciences & Humanities Open

journal homepage: www.sciencedirect.com/journal/social-sciences-and-humanities-open

Regular Article

Democracy's eternal vulnerability: Increasing resilience to disinformation by raising the two components of political truth literacy

Jack Harich [a,*], Montserrat Koloffon Rosas [b]

[a] Thwink.og, 1164 DeLeon Court, Atlanta, GA, 30021, USA
[b] Thwink.org, now at Vrije Universiteit Amsterdam, Amsterdam, Netherlands

ARTICLE INFO

Keywords:
Disinformation
Resilience
Root cause
Democracy
Political truth literacy

ABSTRACT

The vulnerability of democratic systems to political disinformation has become a defining challenge of contemporary governance. Despite considerable attention, theoretical understanding remains undeveloped and solution failure continues. To fill this gap the article develops a starting-point comprehensive theory of resilience to disinformation in democratic systems. Analysis suggests the main root cause of vulnerability to disinformation is low political truth literacy, whose cognitive components are logical truth quotient (LTQ) and appropriate action quotient (AAQ). Both are presently low. Simulation modeling shows that raising them to medium causes the winning political strategy to shift from disinformation to truth-telling. In this state extreme polarization is eliminated, bipartisan consensus becomes the norm, and resilience is achieved. The theory is supported with a proof-of-concept experimental study testing a novel intervention specifically designed to raise LTQ and AAQ. Multiple coordinated solution elements will be required for full resilience.

1. Introduction

The vulnerability of democratic systems to political disinformation has become a defining challenge of contemporary governance. Notable symptoms include climate change denial driven by a well-funded "denial machine" backed by corporate and conservative interests (Dunlap & McCright, 2011, p. 145), the "policy deception" behind the economically disastrous 2016 Brexit vote (Baines et al., 2020), cult-like polarization based on identity politics, state-sponsored disinformation campaigns aimed at electoral manipulation, anti-vaccine movements, economic trickledown theory, intelligent design theory, and scapegoating of immigrants and other groups.

Most ominous is the worldwide surge of far-right populism, which has enabled the ascent of populist authoritarians such as Putin, Modi, Erdogan, Orban, Duterte, and Trump (Scoones et al., 2020, pp. xv, 202). Each has exploited long campaigns of unsubstantiated epistemic claims and conspiracy theories to manufacture a fearful sense of exploitation and persecution of the "noble people" by corrupt elites and other enemies that only a strong leader can save that nation from (Prooijen et al., 2022). The ramifications of political disinformation are "increasingly viewed as a crisis that demands urgent action" (Roozenbeek et al., 2023).

Interventions to counter political disinformation center mainly on fact-checking and media literacy education (Hameleers, 2022), but have proven largely ineffective. While 417 fact-check organizations currently operate in over 100 countries (DRL Staff, 2023), fact-checkers have never solved the core problem of how to get misinformation corrections to the people that need them. "We almost never observe respondents reading a fact-check of a specific claim in a fake news article that they read" (Guess & Nyhan, 2018). Dumitru et al. (2022) find that legal measures have had only slight effect, and that automatic fake news detection and content labeling are too immature or infeasible.

Few long-term, well-implemented cases of media literacy education exist and results are discouraging. The following country-level cases consider only right-wing party behavior. Other data is scarce. The 2022 Media Literary Index (Lessenski, 2022) measures "potential vulnerability" to misinformation for 41 European countries but omits a measure of media literacy itself.

1. Finland, despite compulsory media literacy education from daycare up beginning in the 1970s and ranking number one in the Media Literacy Index, in 2023 saw the populist far-right Finns Party become Finland's second largest party, capturing 20.5 % of the vote (Civicus, 2023).

* Corresponding author.
E-mail address: jack@thwink.org (J. Harich).

https://doi.org/10.1016/j.ssaho.2026.102469
Received 22 June 2025; Received in revised form 8 December 2025; Accepted 9 January 2026
Available online 18 January 2026

2. Sweden, ranking number six with compulsory media literacy education since 1980, in 2022 saw the populist far-right Sweden Democrats become its second largest party, also with 20.5 % of the vote (Erlanger & Anderson, 2022).
3. Denmark, ranking number three with compulsory media literacy education since 1970, saw the populist far-right Danish People's Party rise to 21 % of votes in the 2015 election but then collapse to 2.6 % in the 2022 election. The collapse resulted from the mainstream Social Democrats Party's successful repositioning on immigration, welfare, and redistribution (Etzerpdt & Kohgshoj, 2022; Wikipedians, 2024). Media literacy education failed to prevent the 2015 rise, which could have continued.

Surveying the disinformation field, Bateman and Jackson (2024) concluded that "None of the interventions considered in this report were simultaneously well-studied, very effective, and easy to scale. Rather, the utility of most interventions seems quite uncertain and likely depends on myriad factors that researchers have barely begun to probe."

Thus, no comprehensive disinformation theory exists. To fill this gap, the article develops a starting-point comprehensive theory explaining how democratic systems can be resilient to disinformation attacks. Since this is a first iteration analysis supported by a pilot study of an exceptionally difficult global problem, the theory is far from mature. It offers only a starting point, one designed to be minimally comprehensive but sufficient to productively guide further research toward theory maturation.

The theory comprises a set of interconnected causal hypotheses: Analysis identifies the main root cause of the present lack of resilience as low *political truth literacy*, whose cognitive components are logical truth quotient (LTQ) and appropriate action quotient (AAQ) as defined later. Both are presently low. Simulation shows that raising them to medium causes the winning political strategy to shift from disinformation to truth-telling. As a result, extreme polarization is eliminated, bipartisan consensus becomes the norm, and resilience is achieved. The key hypotheses, that political truth literacy is low and can be raised to at least medium, are tested by a controlled experiment proof-of-concept study.

The research question is thus "What is a starting point comprehensive theory of how democratic systems can be resilient to disinformation attacks." The remainder of the article reviews intervention theory, develops a theory of resilience to disinformation, presents a proof-of-concept experiment to test the theory, and ends with limitations and conclusions.

2. Existing intervention theory

Roozenbeek et al.'s (2023) review illustrates the breadth of current theory by organizing interventions into fourteen categories at the individual and system levels, with most activity at the individual level. Of these, four categories dominate current research and policy attention: fact-checking, critical thinking, media literacy, and inoculation.

2.1. Fact-checking

Fact-checking has been widely implemented. Professional fact-checkers view the rise of misinformation as a threat to "journalism's core mission of helping people to understand public debates and make informed political choices" (Graves, 2016, p. 29). The underlying theory holds that by publishing timely fact-checks on important political claims, fact-checkers can to correct misinformation by providing people with the truth.

The most extensive review of disinformation literature appears in the 130-page report by Bateman and Jackson (2024, p. 3) on *Countering Disinformation Effectively: An Evidence-Based Policy Guide*. The authors found that "fact-checking represents the high-water mark of current knowledge about counter-disinformation measures … [However], numerous knowledge gaps and methodological biases remain even after hundreds of published studies on fact-checking." Their Table 1 rates fact-checking's effectiveness and scalability as modest.

On page 33 Bateman and Jackson explain why fact-checking has had limited impact (italics added):

> [First, it] takes much more time and expertise to produce a fact-check than to generate the false content being debunked. So long as fact-checkers face this structural disadvantage, *fact-checking cannot be a comprehensive solution to disinformation.* … Second, fact-checks require distribution mechanisms capable of competing effectively with the spread of disinformation. … Ideally, fact-checking should occur before or at the same time as the false information is presented. *But this is no easy task*.

2.2. Critical thinking and media literacy

Roozenbeek et al. (2023) classify *critical thinking* and *media literacy* as types of "boosting interventions," which "seek to reduce individual susceptibility to misinformation" by strengthening underlying cognitive competencies**.** One definition of *critical thinking* is "reasonable, reflective thinking that is focused on deciding what to believe or do" (Moore, 2013). The U.S. National Association for Media Literacy Education (NAMLE) defines *media literacy* as "the ability to access, analyze, evaluate, create, and act using all forms of communication," and explicitly includes evaluating disinformation.

In the EU, an "analytical report" prepared for the European Commission (McDougall et al., 2018) posed seven research questions centered on this theme: What media literacy teaching practices would work best to combat disinformation at primary and secondary levels in EU member states? (pp. 17–18) Drawing on 145 publications, the "*guiding analytical framework*" for the report was Hobbs's five essential competencies for media literacy, "as it was developed primarily within and for media education *theory and practice*." (p. 21)

The report also relied on four additional competence frameworks to support "our *analysis* of media literacy education practices" (pp. 19–21) for a total of 44 competencies. Reliance on frameworks arose from a persistent problem: despite decades of research and policy initiatives, there remains "a lack of systematized comparative evidence about 'what works' in media literacy education practices at the classroom level" (p6). Results were a list of 18 Policy Pointers (solution strategies) of a very broad nature oriented toward raising competencies.

2.3. Evaluation

In *Approaches to Social Research*, Singleton and Straits (2005, pp. 18–21, italics added) state the twin objectives of a strong scientific theory are "*to explain the past and present, and predict the future.*" This requires a cohesive set of cause-and-effect propositions that "provide a sense of understanding." They then illustrate how vital both objectives are with a story about discovery of the cure for rickets. Only after the [root] cause of rickets was identified to be a deficiency of exposure to sunlight, necessary for production of calciferol, could rickets be eradicated. This required identification of "the complete causal process." Identifying the problem's causal structure *explained the past and present* (why people had rickets) and successfully *predicted the future* (in terms of intervention design to achieve desired outcomes).

The foundational theory of media literacy education arises from core competencies frameworks, which drive solution design more than anything else in the literature. The theory is simple and implicit: *The higher the competencies, the higher the ability to detect misinformation*.

This is not a scientific theory in the strict sense. It only predicts the future. Missing is the cause-and-effect structure that would *explain the past* (such as why past solutions have largely failed) and *explanatorily predict the future* (such as explaining why particular solution strategies would likely work, which would be the most cost effective, which competencies should be taught, which are more important, and what

levels of competency are required).

In a chapter on Theories in Scientific Research, Bhattachewrjee, 2012, pp. 25–34, italics added) states that:

> A scientific theory … should explain why things happen, rather than just describe or predict.
>
> Note that it is possible to predict events or behaviors using a set of predictors, without necessarily explaining why such events are taking place. … *Prediction requires only correlations*. In contrast, explanations require causations, or understanding of cause-and-effect relationships.

The McDougall et al. report for the EU, and much similar scholarly work, does not develop theory. Rather, it presents a collection of facts, observations, descriptions, correlations, empirical findings, and competencies used to intuitively (rather than analytically) draw conclusions about what solutions will likely work. *A list of competencies is a set of predictors.* The report produced a correlation prediction.

The report relied on a list of competencies due to "a lack of systematized comparative evidence about what works." This evidence will eventually become available. However, a comparative analysis approach, even if done systematically and combined with experimentation, remains a correlation prediction. Due to extreme problem complexity, this will continue to lead to solution designs based on *intuitive* research highly constrained by a *black box model* of understanding, where only inputs, factors, and output are known. What is needed instead is *analytical* research guided by a *glass box model* of understanding of the problem's causal structure.

2.4. Inoculation

The last of Roozenbeek et al.'s four categories to examine is *inoculation*. Here we found a comprehensive scientific theory that uses a cohesive set of cause-and-effect propositions.

Inoculation theory conceptualizes the misinformation problem as a global "infodemic." As Linden (2022) describes, this is *the frame of epidemiology: susceptibility, spread, and immunization.* If analysts can determine why people are susceptible and how misinformation spreads, then effective immunization (inoculation) can be more easily developed.

Linden describes emerging research on immunization as falling into two categories: misinformation correction and prevention. *Corrective solutions* mainly include fact-checks, debunking specific false claims, and general material pointing out the truth. *Preventive solutions* employ *inoculation theory*. By creating a motivation threat against being fooled and then exposing people to a weakened form of the threat, "people can cultivate cognitive resistance against future misinformation."

Inoculation is accomplished with prebunking, "by pre-emptively forewarning and exposing people to severely weakened doses of misinformation, coupled with strong refutations [aka counterarguments]." A limitation is that people must be pre-exposed to all possible types of misinformation they may encounter, which is clearly impossible.

The frame of susceptibility, spread, and immunization is useful. However, the epidemiology analogy has been carried too far by assuming that inoculation is the only way to immunize. Searching for "common building blocks of misinformation" that people can be pre-exposed to is an improvement, but is just as non-scalable as prebunking, since there are thousands and new ones appear daily. More productive is to abandon the constraints of an issue-based inoculation frame and turn to a more suitable frame.

2.5. Going beyond issue-based inoculation with media literacy

In the U.S., media literacy is defined by NAMLE as "the ability to access, analyze, evaluate, create, and act using all forms of communication." The EU defines media literacy as "the ability to access the media, to understand and critically evaluate different aspects of the media and media contexts and to create communications in a variety of contexts" (McDougall et al., 2018, p. 23). The definitions are equivalent since both arise from Hobbs's (2010) framework of the five essential competencies of media literacy. *Here the immunization frame is the correct logical processing of available information.*

How can this be done effectively? Given these continuing antidemocratic trends:

1. The failure of long running media literacy programs in Finland, Sweden, and Denmark.
2. Appearance of a global third wave of authoritarianism, with 75 autocratization episodes since 1976 (Luhrmann & Lindberg, 2019).
3. Election results in Europe, where far-right populist authoritarian candidates won 26 % of the vote in France and 29 % in Austria in 2024. In 2025 the far-right populist authoritarian Alternative for Germany party won 21 % of the vote.
4. 2024 US presidential election results, where far-right populist authoritarian Donald Trump won reelection, a shocking event that signals a new world order and acceleration of the demise of liberal democracy (Kettle, 2024; Kurbjuweit, 2024).

The answer to that question appears unknown.

Bateman and Jackson (2024) concur with this assessment. Because researchers don't know which interventions will work, "Democracies should adopt a portfolio approach to manage *uncertainty*. Policymakers should act like investors, pursuing a diversified mixture of counter-disinformation efforts while *learning and rebalancing* over time." (italics added)

"Learning and rebalancing" describes a trial-and-error process. Hobbs (2020, p. 24) reports trial-and-error was the method used for development of media literacy education practices in the U.S. The same holds for the EU, given lack of causal analysis in the McDougall et al. report and similar research.

High uncertainty and dependence on trial-and-error, combined with the above evaluation, indicate the field lacks a comprehensive theory. Section 4 builds such a theory.

3. Method of research

This is more fully described in Supplementary materials. Here we offer a summary.

3.1. Why we selected root cause analysis (RCA)

Analytical methods commonly used in the social sciences, such as comparative analysis, statistical analysis, path analysis, process tracing, agent-based modeling, expert opinion, case studies, and trial and error via experimentation (hill climbing), perform well on less complex problems. But when faced with problems having high dynamic complexity, these methods lack the required *analytical capacity* because they are not designed to identify causal structure and use that knowledge to generate *glass box models* and causally-informed solution strategy hypotheses, *a task RCA is explicitly designed to do with ease*. For this reason, we selected RCA as the method of analysis.

3.2. How RCA works

RCA works by starting at problem symptoms and asking "WHY does this occur?" until the root causes are found, a generic procedure known as the *Five Whys* (Liker, 2004, pp. 252–256; Ohno, 1988, pp. 17–18). A *root cause* is the deepest cause in a causal chain (or the most basic cause in a feedback loop structure) that can be resolved with practical solutions, without side effects that create other equal or larger problems. *Resolved* means the problem will probably not recur due to that root cause. *RCA* is the systematic practice of finding, resolving, and

preventing the recurrence of the root causes of causal problems. A *causal problem* occurs when problem symptoms have causes, as opposed to non-causal problems like information search and math problems. The golden rule of RCA is: *All causal problems arise from their root causes.*

The basic RCA process is generic and must be wrapped in a process suitable for a particular problem class.

RCA is beginning to make inroads into social science, such as Wagner's (2014) eight-step RCA process for public policy pedagogy, finding the root causes of health inequity (Weinstein et al., 2017), finding the root causes of racial disparity (Conley et al., 2024), finding the root causes of biodiversity loss (Wood et al., 2000), general use of RCA in health care systems (Percarpio et al., 2008), and use of RCA for evaluating public program improvement (Coskun et al., 2012). These studies demonstrate RCA is has become an accepted practice in the social sciences. However, no RCA-based method suitable for difficult large-scale social problems was found. We were thus compelled to develop one. The result was social force diagrams, a standard fill-in-the-blanks template that guides RCA.

While RCA is still in its infancy in the social sciences, some evidence of effectiveness is beginning to emerge. Most organization uses of RCA are to incrementally improve existing processes, rather than to solve difficult large-scale social problems. In health care, of the 11 articles that

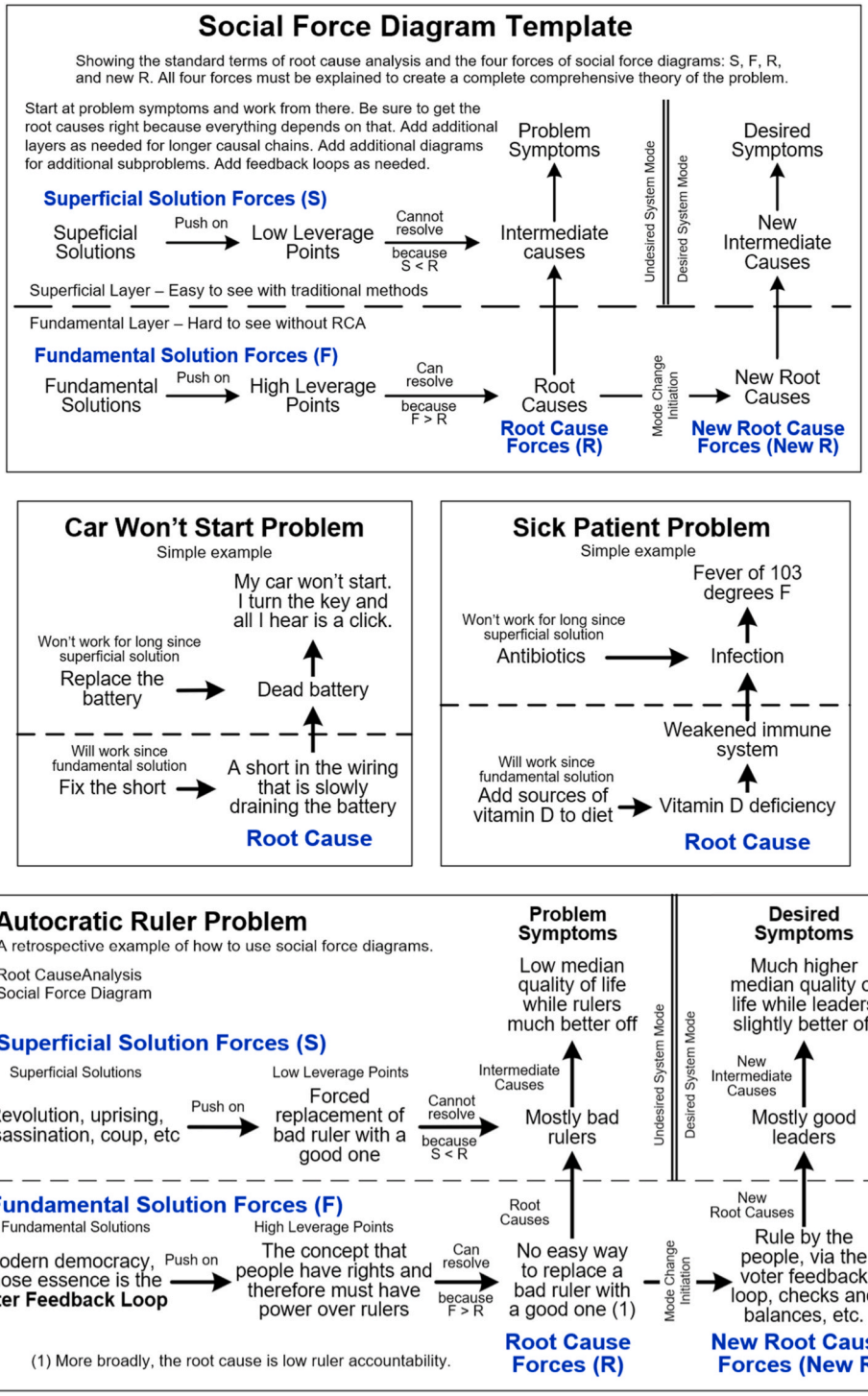

Fig. 1. Social force diagram template, two simple examples, and a retrospective example.

that measured RCA effectiveness, all reported improvement of patient safety (Percarpio et al., 2008). In basic research, the goal is understanding why a phenomenon occurs, with no intent to solve a problem. Using RCA, Byers et al. (2012) analyzed why merchants using Groupon offers saw sharp declines in their Yelp ratings scores. The root cause was counter intuitive. Use of Groupon offers *increased* Yelp reviews and review scores by real users. This *decreased* average Yelp scores, because so many Yelp reviews are fakes with artificially high scores.

3.3. Social force diagrams

RCA is based on the principle of discovering the causal structure essential to solving a causal problem and ignoring everything else. The principle is embodied in the concept of essential causal structure (ECS), which is the minimum amount of causal structure that contains the nodes (variables), relationships (arrows), causal chains, and feedback loops needed to analyze and solve a causal problem. To be complete, this must include problem symptoms, intermediate causes, and root causes. Explicitly identifying and testing ECS is the core of systems thinking and the core of building a strong scientific theory. The ECS paradigm can lead to surprisingly simple but useful causal structures. The basic structural shape of ECS is:

Root Causes → Intermediate Causes → Problem Symptoms

RCA is generic. Thus, so is ECS. For the class of difficult large-scale social problems, the above three-node structure must be expanded considerably. The result is social force diagrams (Fig. 1). These are organized into two main layers:

1. The **superficial layer** of the problem, where intermediate causes are so easy to see they are erroneously assumed to be root causes, and
2. The deeper **fundamental layer**, where by understanding the problem's deeper structure its well-hidden root causes may be found. Without analysis of the fundamental layer, difficult problems tend to remain stuck in the superficial layer for a long time or indefinitely.

The two layers allow avoidance of the *Superficial Solutions Trap*. The trap occurs when problem solvers unknowingly assume intermediate causes are root causes, and then develop solutions based on that false assumption. This leads to solutions directed toward intermediate causes rather than root causes. Superficial solutions can *never* resolve root causes because root cause forces exert a greater force on intermediate causes than superficial solutions, regardless of how well funded, managed, or promoted those solutions are.

3.4. The four forces of social force diagrams

Social force diagrams focus on understanding four key forces: S, F, R, and new R. Superficial solutions (force S) fail because *force S is always less than root cause forces* (force R), indicated by **S < R**. By contrast, if fundamental solutions (force F) are properly designed (especially their impact on feedback loop structure), *force F can exceed force R*, indicated by **F > R**. This leads to a systemic mode change, during which the old R is replaced by a new R, and the problem is solved. The new R must be engineered to be strong enough and self-managing enough to permanently hold the system in the solved mode, due to the way force F fundamentally changes critical feedback loop structure and loop dominance.

If analysis shows no F > R exists (no resolvable root cause is found), the problem is unsolvable as defined. In this case problem definition (problem symptoms) can sometimes be relaxed to make the problem solvable, such as raising the maximum allowable global temperature rise for climate change to make that problem solvable. Otherwise, the problem should be declared unsolvable.

Once all four forces are understood and key assumptions have been tested by measurement or experimentation, the analyst has a sufficiently complete scientific theory of the problem. Each force provides an explanatory tenet of the theory. This gives *the four requirements for a comprehensive theory of a difficult large-scale social problem.* The theory must identify the four forces and explain them in this manner:

1. **Social Force S** – Why past solutions failed (because S < R).
2. **Social Force R** – Why the problem occurs (because R is unresolved).
3. **Social Force F** – Why fundamental solutions can be expected to succeed in causing the desired mode change (because F > R).
4. **New Social Force R** – Why the mode change will be relatively permanent (because the new R contains self-sustaining feedback loops).

4. Theory of resilience to disinformation

4.1. Analysis social force diagram

Fig. 2 summarizes analysis results. To read the diagram, start at Problem Symptoms and follow the arrows backward.

The diagram was built incrementally by starting with Problem Symptoms and working downward from there. We had strong clues to work with: a collection of superficial solutions that had failed, despite repeated, long-term attempts to modify them to improve success. None achieved more than modest success, due to the Superficial Solutions Trap. All the superficial solutions in Fig. 2 fell into this trap.

The first step of analysis is defining the problem in terms of its symptoms. Scholars are concerned about political disinformation because if too many citizens are fooled, this causes problem symptoms of *election of politicians not working for the democratic common good*. For example, a politician who denies climate change will resist climate-policy solutions, a populist-authoritarian will promote democratic erosion, an anti-vaccine politician will undermine public health through anti-vaccine policies, and so on.

4.2. Analyzing the superficial layer

WHY do these problem symptoms occur? Because of *belief in damaging falsehoods like climate change denial, populist authoritarianism, anti-vaccine*. The low leverage point strategy to solve this problem is *misinformation correction by providing people with the truth.* Common solutions are *fact-checks, debunking, articles, social media posts, news, content labels, etc. pointing out the truth*. These generally fail because misinformation correction is a low leverage point. Once a person believes a lie, they seldom encounter the appropriate correction and are heavily influenced by motivated reasoning (Lodge & Taber, 2013), which reduces correction effectiveness.

Next, WHY does *belief in damaging falsehoods* occur? The answer is mainly because of *successful political deception*. This presently works so well that *politicians destructively compete to see who can deceive the most voters to vote for them.*

In a democracy, powerful special interests (economic elites and powerful politicians) are by definition a small minority. This raises a pivotal question: How can a *small minority* working for the uncommon good persuade a *majority* working for the common good to vote against their own best interests?

In a democracy, the main ways to do this are by force, threats, rigged elections, voter suppression, favoritism, bribes, or deception. Force, threats, and rigged elections are illegal. Voter suppression is mostly illegal. Favoritism doesn't work on large populations, since there are not enough favors (like jobs or contracts) to dole out. Bribes are inefficient, as even the rich lack the resources to bribe millions of voters. This leaves deception as the main preferred strategy and explains why *successful political deception* is required.

Jeremy Bentham, the father of utilitarianism, reached the same conclusion in 1824:

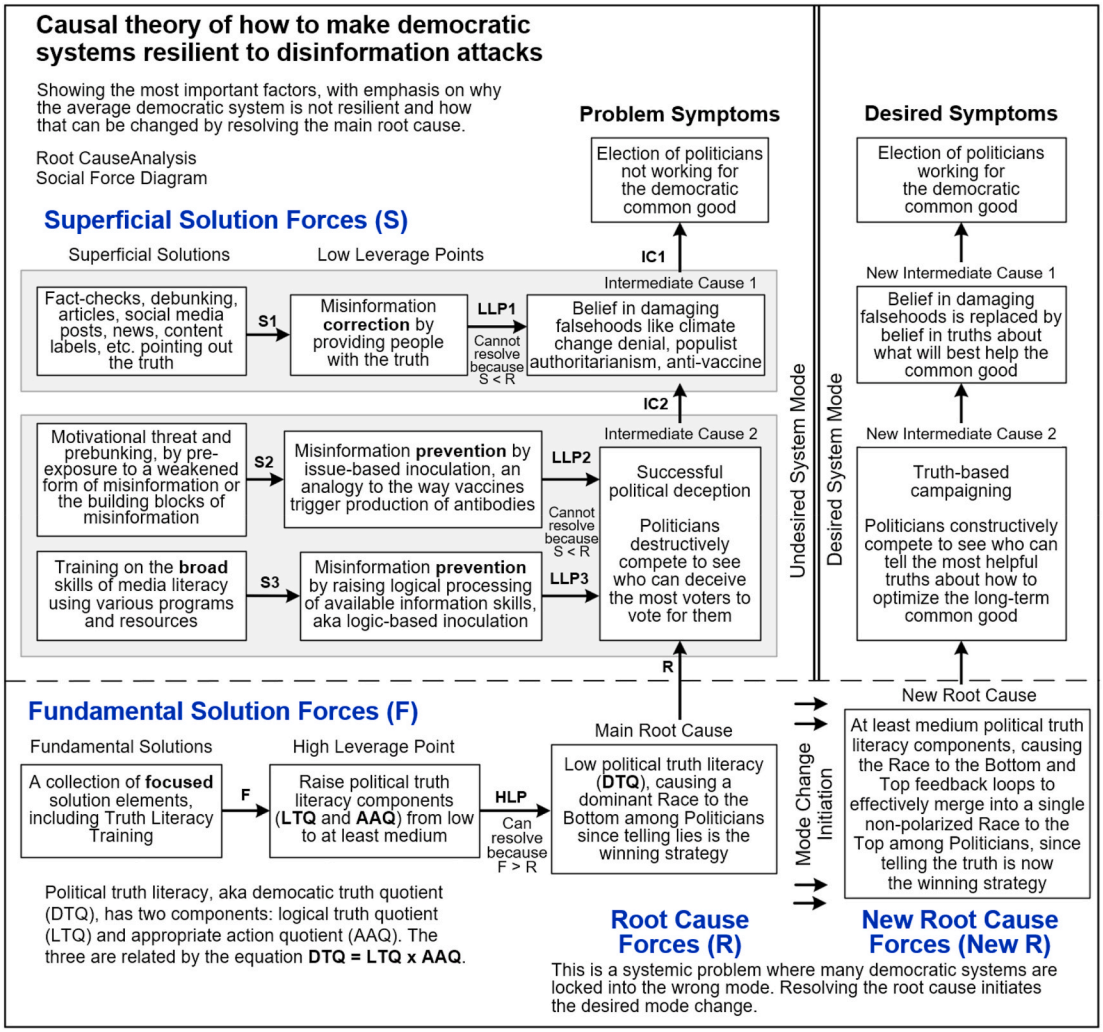

Fig. 2. Social force diagram of the theory.

> … it is impossible by fair reasoning … to justify the sacrifice of the interests of the many to the interests of the few …. It follows that for effecting this purpose they must have recourse to every kind of fallacy, and address themselves, when occasion requires it, to the passions, the prejudices, and the ignorance of mankind. (Larrabee, 1952, p. xxi)

Democracy's eternal vulnerability to political deception has existed since the dawn of democracy. When making the case that deceit by elite politicians posed a profound threat to Athens's democratic process, Demosthenes reached this conclusion in the fourth century BC:

> A man can do you no greater injustice that to tell lies. In a political system based on speeches, how can it be safely administered if the speeches are not true? (Hesk, 2000, pp. 1–2)

Two main low leverage point strategies have been used to prevent *successful political deception*. The first is *misinformation prevention by issue-based inoculation, an analogy to the way vaccines trigger production of antibodies*. Linden (2022) describes how this is accomplished by *motivational threat and prebunking, by pre-exposure to a weakened form of misinformation or the building blocks of misinformation*. However, a limitation is that people must be pre-exposed to all possible types of misinformation they may encounter, which is clearly impossible. Linden acknowledges this and reports that:

> Instead, scholars have started to identify the *common building blocks of misinformation* more generally, including techniques such as impersonating fake experts and doctors, manipulating people's emotions with fear appeals, and the use of conspiracy theories. Research has found that people can be inoculated against these underlying strategies and, as a result, become relatively more immune to a whole range of misinformation that makes use of these tactics.

This describes a promising alternative to issue-based inoculation, known as technique or logic-based inoculation, which "confers resistance against manipulation strategies or tactics such as logical fallacies, emotional manipulation, or conspiracy theories" (Roozenbeek et al., 2023). This leads to the second low leverage point of *misinformation prevention by raising logical processing of available information skills, aka logic-based inoculation*. Logic-based inoculation research remains formative and fragmented across exploratory studies on a wide range of potential approaches. The longest and most common form of logic-based inoculation has been *training on the broad skills of media literacy using various programs and resources*. As described in the Introduction, these interventions have failed to have the intended effect.

The three groups of superficial solutions are working so poorly we are now living in the post-truth age of politics, "in which lies and distortions carry as much weight as facts" (Puddington, 2017, p. 57).

4.3. Finding the main root cause

Next, WHY does *successful political deception* occur?
Could it be low levels of literacy or education? We ruled that out

because successful political deception, particularly far-right populism, occurs in nations with high levels of functional literacy (the ability to apply reading and reasoning skills in real-world contexts), like the US and EU, as well as low levels, like South America, as measured by the OECD's PISA test of 81 countries (OECD, 2023, p. 97). Missing is *the specific component of literacy* that would prevent successful political deception, which this paper identifies as political truth literacy.

Could it be due to cognitive flaws like confirmation bias or the *processing fluency effect*, where "the more a claim is repeated, the more familiar it becomes … the brain uses this fluency as a signal for truth" (Linden, 2022). We ruled this out because these flaws are relatively constant, while outbreaks of successful political deception come and go, such as the three waves of autocratization beginning in 1926, 1961, and 1994 (Luhrmann & Lindberg, 2019).

Could it be high levels of economic inequality or insecurity? Many studies have found a correlation between these conditions and the rise of far-right populism, e.g. (Gidron & Hall, 2017; Rooduijn & Burgoon, 2018). But correlation is not cause. We interpret this relationship to indicate an *exploitable grievance*. This causes higher susceptibility to well-targeted fallacious appeals, such as a far-right populist claiming "your money is being taken by corrupt elites and the filthy rich, and only a strong leader like me can fix that."

Finally, could it be due to political or cultural beliefs that predispose people to accept clever fallacious appeals? Yes, to some extent. Such beliefs often contain distortions, motivated reasoning biases, or loyalty to group identity rather than to truth and the common good. Rather than exploitable grievances, they are *exploitable false beliefs* that can be magnified into major misconceptions, such as hatred of everyone not in your identity group (extreme identity politics) or of a particular group like immigrants (scapegoating). Lodge and Taber (2013, p. 150) explain how "citizens are often partisan in their political processing, motivated more by their desire to maintain prior beliefs and feelings than by their desire to make accurate or otherwise optimal decisions," which is a highly exploitable state.

A deeper question remains. Political and cultural beliefs are learned from the political parties, news sources, books, churches, cultures, family members, teachers, friends, etc. a person is exposed to. But why would a person make an error and believe false statements, regardless of the source or prior beliefs, and develop erroneous beliefs? *Why exactly does political deception succeed?*

Answering that question at the root cause level blocked progress for years and was so difficult it required construction of a system dynamics simulation model (Fig. 3).

The backbone of the model is the two opposing feedback loops dueling for the same Uncommitted Supporters. Race to the Bottom politicians, the right, use deception to gain supporters, while Race to the Top politicians, the left, use the truth. Currently the Race to the Bottom is the dominant loop in many countries most of the time because *undetected false memes* (successful misinformation) are high.

The model uses the concept of memes (Dawkins, 1976). A meme is copied information capable of affecting behavior, such as a fact or opinion. In the model a meme is a statement that is true or false, regardless of its form, such as conversation, social media, print media, image, video, etc. Dawkins introduced the concept of memes and memetic infection. This was taken up by others, such as in Brodie's (1996, pp. xvi–xvii) *Virus of the Mind: The New Science of the Meme:*

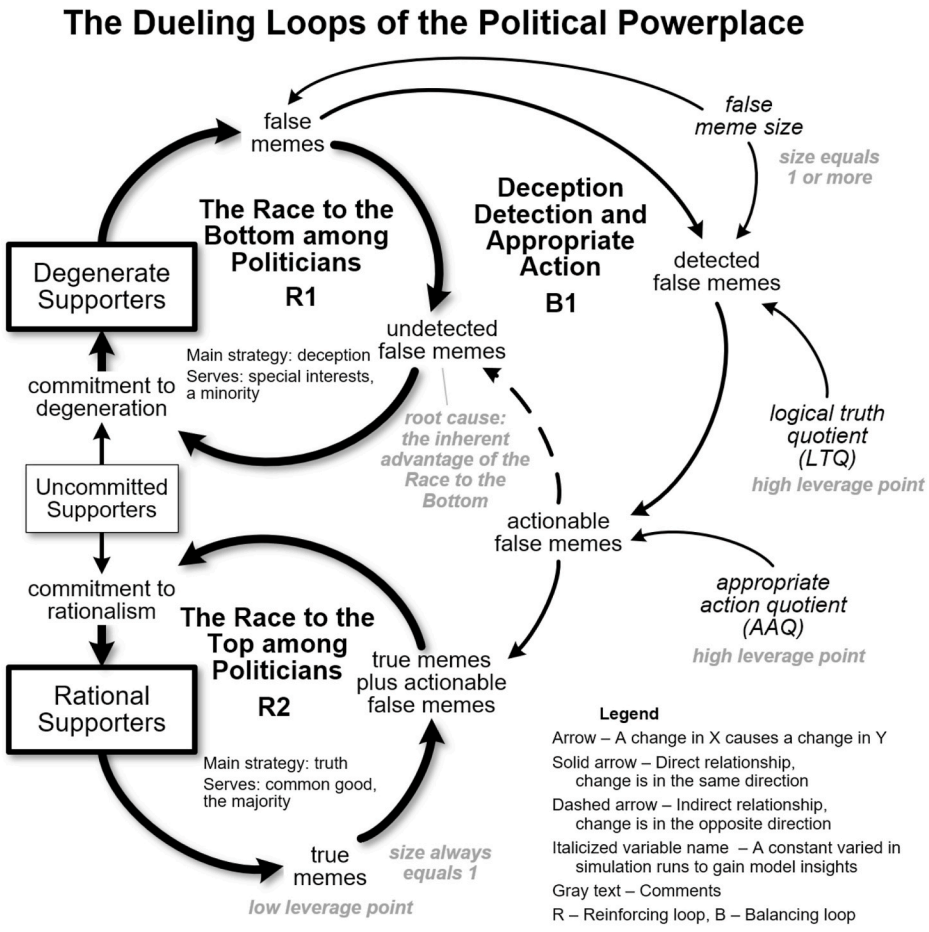

Fig. 3. Feedback loop structure explaining how political deception works. This is a high-level diagram of the simulation model. See supplementary materials for detailed model description.

> [A meme] is the basic building block of culture, in the same way the gene is the basic building block of life. … Once created, a virus of the mind [a meme] gains a life independent of its creator and evolves quickly to infect as many people as possible.

Thus, the spread of misinformation can be realistically modeled as the process of memetic infection, as memes spread from one mind to another. Our model is similar to epidemiology's SIR model (Sterman, 2000, pp. 303–304) in that it uses infection rates, average duration of infection, and susceptible and infected populations, but incorporates many additional features necessary for modeling root cause behavior.

The key model insight is that the size (and hence the attractive power) of a lie (*false memes* on the model) can be inflated, while the size of the truth (*true memes*) cannot. From a mathematical perspective, the size of a falsehood can be inflated to be a different, more attractive (false) reality by saying that $2 + 2 = 5$, or 7, or even 27. But the size of the truth is always 1 since there is only one true reality. Its attractive power can never be inflated by saying anything more than $2 + 2 = 4$. Inflation is used to create fear when there is nothing to fear, doubt when there is nothing to doubt, the false promise of I can do so-and-so for you when I really cannot, a large flaw in one's opponent when there is only a small flaw or no flaw, etc.

This insight leads to identification of the main root cause: the inherent asymmetric advantage of the Race to the Bottom, represented on the model by *undetected false memes*. The advantage exists because the opposing loop, the Race to the Top, has no corresponding variable because there are no inflated *true memes* to detect. This reveals a second form of the asymmetric nature of information warfare. The first form, that "autocracies [can] penetrate the information space of democracies without facing an equivalent response" (Liagusha & Iarovyi, 2025), is far more obvious.

For simplicity we usually say the main root cause is *low political truth literacy*. This causes a dominant Race to the Bottom most of the time since telling lies is the winning strategy. This appears to be a root cause rather than an intermediate cause because there is no deeper root cause and practical solutions for resolving it exist.

4.4. Evidence of universal reliance on political deception by right-wing politicians

If *low political truth literacy* is the main root cause, one would expect universal reliance on political deception by Race to the Bottom parties and politicians. Evidence of this is considerable:

1. Freelon et al. (2020) found that "in the US and throughout the industrialized West … available evidence suggests that the right has invested far more than the left in misinformation and conspiracy theories as core components of its activist repertoire …."
2. An examination of political behavior in 169 countries in 1759 elections found four key characteristics of right-wing autocracy (Luhrmann et al., 2023). All require deception to implement.
3. In the U.S. a well-financed "Right-Wing Propaganda Machine" has dominated political debate for decades (Conason, 2004).
4. Dunlap & Jacques, 2013 found that 92 % of books denying climate change and other environmental sustainability problems originated from conservative think tanks.
5. In an effort to understand the effect of propaganda on politics, Benkler et al. (2018, pp. 77, 79) analyzed four million messages in the US using their Media Cloud platform. America's political spectrum has evolved into two opposing feedback loops: a right-wing "propaganda feedback loop" where politicians "compete on identity confirmation" regardless of the truth, versus a centrist/left-wing "reality-check" loop that follows "institutionalized truth-seeking norms" where politicians "compete on truth quality and the scoop". The two loops correspond exactly to the Race to the Bottom and Top loops.

By logical implication, if successful political deception is widespread, then political truth literacy must be low. If it were not, political deception would usually fail. This logical inference leads to hypothesis H1 (that political truth literacy is currently low), which is tested in the Truth Literacy Training study.

4.5. A two-step cognitive model of political truth literacy

Truth literacy is the ability to tell truth from deception, to be able to "read" the truth. A *political claim* is any claim meant to influence voters. The analysis uses three important variables to model the cognitive behavior of political truth literacy, or any form of decision making involving potentially deceptive claims. All range from zero to 100 %:

1. **LTQ** (*logical truth quotient*) is the ability to logically tell whether a political claim is true or false.
2. **AAQ** (*appropriate action quotient*) is the ability to take appropriate action, given the perceived truth (using LTQ) of a political claim.
3. **DTQ** (*democratic truth quotient*, aka political truth literacy) is the ability to take informed and logically correct action within a democratic system, particularly voting, given a possibly deceptive political claim.

A person's DTQ uses the two-step process of (1) determine the truth (LTQ) and then (2) take action given that perceived truth (AAQ). For example, "I can see that statement is false because it uses the false dilemma fallacy." And then "Now I need to vote against that politician because they cannot be trusted to tell me the truth." Because of this two-step process, DTQ = LTQ x AAQ, though DTQ is not in the model.

For example, suppose a person's LTQ and AAQ are low at 10 % and 20 %. They will detect 10 % of lies using their LTQ. Then, using their AAQ, they will use 20 % of the 10 % to take appropriate action. $10\ \% \times 20\ \% = 2\ \%$. They will respond correctly 2 % of the time to misinformation. The Race to the Bottom will dominate.

Suppose their LTQ and AAQ are both medium at 50 %. Now they do much better and detect 50 % of lies. Since $50\ \% \times 50\ \% = 25\ \%$, they will respond correctly 25 % of the time, an order of magnitude better. While they are still fooled most of the time, the simulation model shows that enough lies are detected to make the Race to the Top dominant.

4.6. Resolving the root cause

The single high leverage point in the social force diagram, *raise political truth literacy components from low to at least medium*, becomes two high leverage points in the simulation model: LTQ and AAQ. Presently both are low, which gives the Race to the Bottom its asymmetric advantage and allows successful deception to occur. Because both are low, in simulation runs optimum *false meme size* is much more than one. Many large lies occur, causing Race to the Bottom dominance. As LTQ and AAQ are raised, more lies are detected. In an effort to maximize the number of Degenerate Supporters, optimum *false meme size* falls to make lies harder to spot. Lies get smaller.

As LTQ and AAQ are gradually raised to 50 %, optimum *false meme size* falls all the way to one, which is no lying at all. Supporters on the right have moved from an extreme right position (the far-right) based on a false ideology to a truth-based moderate position, one so moderate that like the left, the right also pursues the common good. Neither side has an advantage since lying ceases. *The two loops effectively merge into a single non-polarized Race to the Top* with approximately equal percentages of degenerates and rationalists. There will be disagreement between the left and right, which is normal in politics, but it will be constructive and small enough for the system to solve common good problems. Extreme polarization is eliminated, bipartisan consensus becomes the norm, and resilience to disinformation is achieved. The system has undergone a fundamental mode change from undesired to desired symptoms.

We are not saying all polarization will be eliminated, only extreme

polarization. This model-based prediction is an example of what Sterman (2000, pp. 5–23) calls the dynamic complexity of complex social systems and their counterintuitive, unpredictable behavior—until their correct high leverage points are known.

4.7. Filling the gap in media literacy education theory

Of all the solutions studied, media literacy education came the closest to being able to push on the high leverage point of *raise political truth literacy components (LTQ and AAQ) from low to at least medium*.

Political truth literacy (DTQ) is defined as the ability to tell truth from falsehood in statements meant to influence voters and to make logically correct voting decisions based on that information. *DTQ is a very small subset of media literacy*, broadly defined in the US by NAMLE as "the ability to access, analyze, evaluate, create, and act using all forms of communication." The definition is extended by the Six Core Principles (Potter, 2022, p. 32) which broaden the definition still further. The EU uses a similar definition. The pattern is media literacy definitions of what should be taught tend to be all-encompassing and aspirational.

The strategic reason present media literacy education efforts fail to prevent susceptibility to political disinformation is they are too broad. They do not properly focus on the very narrow set of skills (LTQ and AAQ) required to resolve the main root cause. In the study presented later, we demonstrate how these skills can be taught in a low-cost practical manner.

4.8. Filling the gap in motivated reasoning theory

Building on extensive experimental evidence, Lodge and Taber (2013, pp. 24–25, italics added) explain how:

> Our model asserts that *motivated reasoning*—the systematic biasing of judgements in favor of *automatically activated*, affectively *congruent beliefs and feelings*—is built into the basic architecture and information processing mechanisms of the brain. … [Because of this,] spontaneous actions are difficult to correct, even when people are encouraged to stop, think, deliberate, or actively try to work their way through a problem.

Voter reasoning is motivated and largely unconscious. When faced with a deceptive political appeal, people default to motivated reasoning and are easily fooled by any appeal aligning with preexisting "congruent beliefs and feelings." This explains why Lodge and Taber concluded that "we see no obvious resolution of the dilemma" (p27) of the influence of motivated reasoning in politics.

"Automatically activated" means the error of motivated reasoning occurs subconsciously. This is the mechanistic barrier that current theory fails to surmount.

We hypothesize the error occurs because the incorrect reasoning rules used are informally assimilated, rather than the correct rules being formally taught. The result is today's voters are functionally illiterate in discerning political truth.

In the study presented in the next section, we demonstrate how correct rules can be formally taught. We further hypothesize that LTQ and AAQ, as the components of political truth literacy, behave in a similar manner to components of reading and writing literacy in that they are complex, heavily used in daily thinking, and must be taught.

Once a person has been taught and becomes politically truth literate, they no longer make the error of motivated reasoning because they are now following a different rule set. In training and early use, this occurs with some reasoning effort at the conscious level. Later after much practice, rule set use moves to the subconscious level where most daily reasoning occurs.

We support Lodge and Taber's claim that "Once attitudes have become crystallized, persuasion is difficult" (p152). We thus do not propose that solution elements like Truth Literacy Training can undo the strong synaptic connections that lock some voters into falsehood-based highly polarized states, such as Trump's MAGA supporters. Less-polarized or non-polarized voters, whose political beliefs have not overly crystallized, can benefit from the training. This would include swing voters and the young.

5. The Truth Literacy Training study

The analysis identified deficiencies in two major disinformation theories. This section demonstrates:

1. How the deficiency of *media literacy education theory*, too broad a goal due to the very broad skillset of media literacy as presently defined, can be overcome by tightly focusing on just LTQ and AAQ skills.
2. How the deficiency of *motivated reasoning theory*, errors due to informal assimilation of incorrect decision-making rules, can be overcome by formal teaching of correct rules for processing political disinformation.

A collection of solution elements was designed to push on the high leverage point of *raise political truth literacy components from low to at least medium*. The most promising one to develop first was Truth Literacy Training, since it requires the least amount of work to develop, test, and implement for the highest impact, and is the core solution element.

The key output of the analysis was two hypotheses:

H1. The main root cause of low political truth literacy exists.

H2. The root cause can be resolved in a practical manner.

To make a first pass at testing these hypotheses, we developed and performed the proof-of-concept Truth Literacy Training study.

5.1. Study design

The complex user interface required for online training required custom software and our own database. The final format consisted of a long questionnaire supplemented by training materials (Fig. 4).

The questionnaire consisted of five sections as seen in the Outline panel in Fig. 4. The only section that varied per group was the Review Section, which varied the type of training as described below.

The subject in Fig. 4 has completed the LTQ and AAQ training, done in the Getting Started and Review Section in the left panel. They have just answered three questions concerning a statement about a Trade Agreement Treaty. All three answers are correct. All statements follow a similar format, with the claim bolded. All questions are unrelated to each other and stand alone.

The Personal Truth Test is shown on the right panel. Notes are the two vote training rules. The subject scrolls up to see the rest of the Reference Material. This includes a catalog of fallacies they were trained on and a graphic of the Strong Evidence Rule (Fig. 5). Using the left panel, subjects can navigate anywhere in the questionnaire to review their decisions or training material. Checks indicate a completed item. Decision answers can be easily reviewed and changed.

Using a Prolific.com online panel filtered to US subjects aged 22 to 99, the study was run on 93 subjects randomly assigned to three groups. A 22-year minimum was used since pretesting showed subjects younger than 22 did not take the questionnaire seriously, raced through, and had unreliable, wildly varying answers with lots of guessing (random answers). Resultant demographics were age range 22–51, average age 31, 49 % male. Educational levels were 34 % high school, 55 % college degree, 10 % PhD. All were told this is a decision-making study for the purpose of improving the health of democracy. The three treatment groups were:

1. **Control group** – This group received training on the neutral topic of how democracy works.

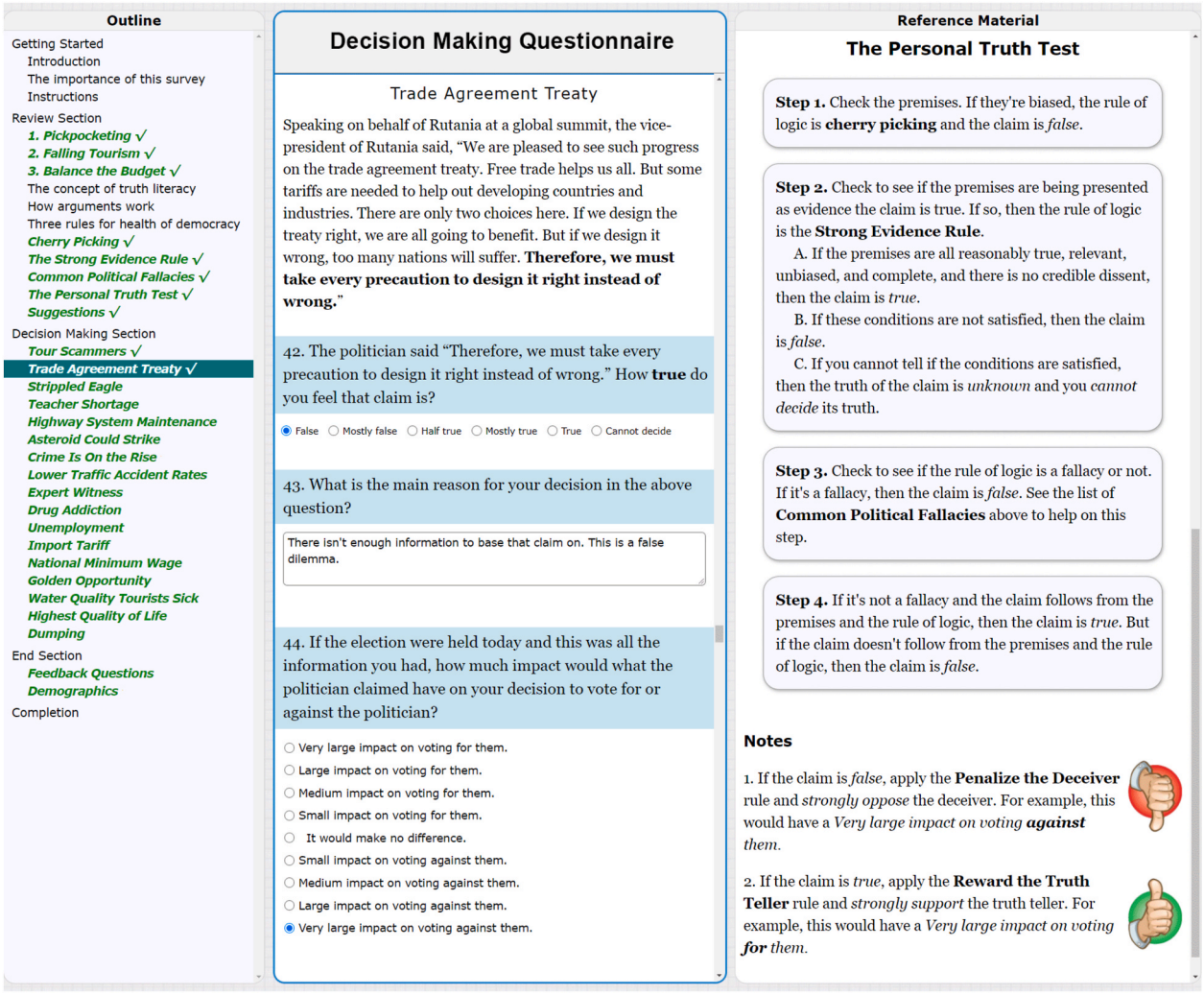

Fig. 4. User interface for the online Truth Literacy Training study, group 3. This runs in any popular browser.

2. **Claim training** (LTQ) – This group received training on how to tell if a political claim (embedded in a political statement, such as the one in Fig. 4) was true or false, by spotting the pattern of fallacy or non-fallacy used and using the Personal Truth Test, which includes the Strong Evidence Rule. The strategy is high-speed pattern recognition to drive the Personal Truth Test. Humans are excellent at pattern recognition, if they know what to look for. "… the ways by which we are deceived are consistent and not so hard to recognize" (Jackson & Jamieson, 2007, p. 6).
3. **Claim and vote training** (LTQ and AAQ) – This group received the same training as group 2 plus training on how to vote correctly (given the perceived truth of a claim) by applying two rules: Reward the Truth Teller and Penalize the Deceiver (See Fig. 4 Notes section for these rules.). Total time for group 3 averaged 87 min, of which about 1 h was training. Group 3 training used 37 questions.

At one point in training development, we found many subjects were not taking the training seriously, resulting in widely varying and mostly low scores.

The problem was solved by discovery of an insightful set of experiments on "dispelling the illusions of invulnerability" to deceptive persuasion (Sagarin et al., 2002). The study found it was not enough to expose subjects to deceptive statements and explanations of why they were deceptive. This failed to work because "our participants' sense of unique invulnerability to deceptive ads left them unmotivated to use defenses against such ads." This illusion of invulnerability caused subjects to believe they were not susceptible to deception, with the result that "they did not resist the ads containing illegitimate [deceptive] authorities more effectively than did controls." This was corrected by "demonstrating in an undeniable fashion that participants can be fooled by ads containing counterfeit authorities."

To dramatically demonstrate to subjects they are not invulnerable to deception, we changed the initial part of the training. After subjects in groups 2 and 3 answer questions for the first three statements in the Review Section of Fig. 4 and before any training has occurred, they read an educational item on *The concept of truth literacy*. There they are shown their own answers for the first three statements (the first two are almost always wrong) versus the correct answers. The item then says:

> If you got all the answers right, congratulations. However, here's how other people did. In a past survey with 34 participants, none got the answer to the first question right. Three got the answer to the second question right. On the third question 19 people got the answer right.
>
> Why are the first two questions so hard? It's because they use clever forms of deception, which makes it terribly difficult to determine how true the claims are.
>
> The reason so many people got the third question right is it's not deceptive. Generally, it's much easier to spot the truth as opposed to

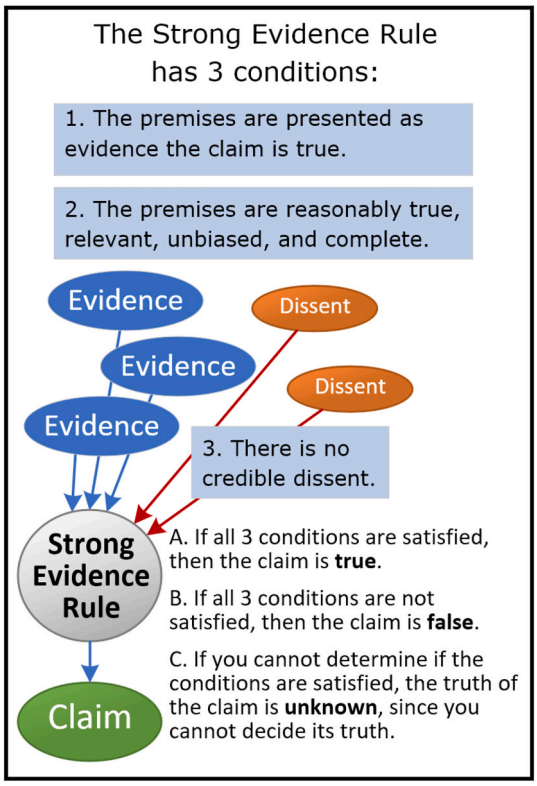

Fig. 5. The strong evidence rule.

deception, because we are so used to processing true statements from people we talk to, books we read, and so on. …

Fortunately, there's a solution to this problem.

Here's the solution. The reason citizens are so easily fooled by deceptive statements is **low truth literacy**. The average person has never been trained in telling truth from deception, so their truth literacy is low. Because it's low, they are unable to reliably tell truth from deception.

For example, the average person is unable to instantly see that the claims in the first two statements are false, because they both use the **cherry picking** fallacy.

Truth literacy is the ability to tell truth from deception. Universal truth literacy is just as important to the health of democracy as reading literacy, because if people cannot "read" the truth they are blind to what the truth really is. They are easily controlled by any politician who uses deception to hoodwink the masses into supporting him and his positions.

In the Truth Literacy Training that follows you are going to learn two things:

1. How to spot the **cherry picking** fallacy, a form of deception.
2. How to spot the **strong evidence** non-fallacy, a form of telling the truth.

This and the material that follows triggers realization they are vulnerable to deception. From this point on, almost all take the questionnaire seriously. From the viewpoint of the elaboration likelihood model of persuasion (Dillard & Shen, 2013, pp. 137–149), elaboration *motivation* has increased from low to high. Subsequent training increases their elaboration *ability*, with the result that when training is complete, most deceptive persuasion attempts will be processed (elaborated) correctly and they will not be fooled.

There is a 5-min break after training for all groups, necessary to avoid fatigue and loss of interest on such a long questionnaire.

In the questionnaire, non-hot statements were presented in random order. Each statement was followed by three questions: (1) the truth question, (2) an open-ended question designed to maintain cognitive motivation and give us feedback, and (3) the vote question. The fictious country of "Rutania" was used in statements to create interest and political realism without the hot bias a real country would have provoked. Deceptive statements contained six fallacies we found common in political appeals: cherry picking, ad hominem attack, appeal to emotion, strawman, false dilemma, and false fact lie, plus flawed application of the Strong Evidence Rule.

5.2. *Study results*

Results are shown in Fig. 6. Rather than a 95 % confidence interval calculation using the central limit theorem, bootstrapping resampling with replacement was used (Banjanovic & Osborne, 2016), due to the data being far from the non-skewed normal distribution assumption required for the central limit theorem.

Results were positive. Average DTQ for the control group was low, 2 %, with a 95 % confidence range of 1% to 4%. *This offers* preliminary proof-of-concept support for H1, *that the root cause of low political truth literacy exists.* Average DTQ for the fully trained group (group 3) was 67 %, a 65-point rise. Full training averaged 1 h. *This suggests that Truth Literacy Training* can successfully push *on the high leverage point of raise political truth literacy components from low to at least medium.* This confirms the potential feasibility of H2, that the main root cause can be resolved.

The literature offers some support for a generally low level of DTQ. A study on the teaching of critical literacy in UK schools found 2 % of children and young people, and 4 % of adults could successfully determine the truth or falsity of six short news stories correctly (Powell, 2023).

The follow up study 26 days later (using different statements) found LTQ and DTQ for group 3 had declined from 76 % to 66 % and 67 %–60 %, 10-point and 7-point falls. After an average of 30 min of refresh training, LTQ and DTQ for group 3 rose to 75 % and 70 %, indicating regular refresh training of some type can work and will be required. Or it may be that like reading and writing literacy, once political truth literacy matures and becomes the reasoning default and is exercised often enough, little decline will occur. Long term it may even rise, if reinforced and strengthened by regular exposure to Truth Literacy Training topics in news media, such as stories centered on use of a particular fallacy.

As predicted, LTQ training alone (group 2) does not raise DTQ significantly. AAQ training is also required (group 3). Still, the 6 % DTQ for group 2 was an astonishing discovery. *We expected it to be low, but not that low.* Even if a person has been trained on how to tell whether a political claim is true or false, they are unable to translate the truth or falsity of a claim into correct action. Instead, they choose all sorts of answers for the vote question.

This indicates the average voter currently does not fully *penalize politicians they know to be deceptive.* This behavior is required for democratic governments to work in the best interests of voters. We suspect the main reason for this dysfunctional behavior is that hardly anyone has received the equivalent of Truth Literacy Training and in particular vote training, which is amazingly simple. Vote training consists of following the two simple rules described in the lower right of Fig. 4.

For the first questionnaire, treatments groups 1, 2, and 3 had N = 30, 30, and 33. Small N was chosen due the pilot nature of the study and the high cost of reimbursement per subject, as total time averaged 2.5 h per subject across both questionnaires during pretesting. Small N is typical for pilot studies, especially when large effects are anticipated (Bell et al., 2018). Final pretesting with N = 6 for group 3 found a 50 to 70

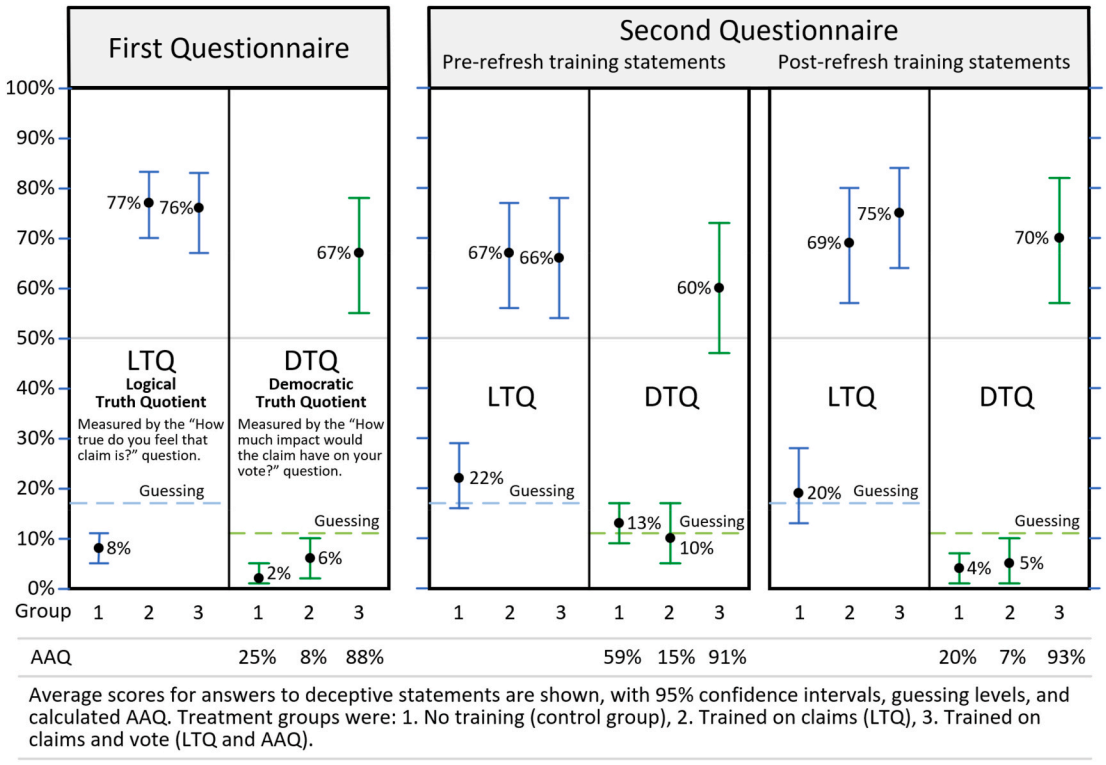

Fig. 6. Results of the Truth Literacy Training study.

percentage-point DTQ increase. Using a sample size calculator (www.powercalc.ca) for anticipated DTQ of 5 % and 55 % for groups 1 and 3 (a 50-point rise), estimated standard deviation 50 percentage points, desired two-sided significance .05, and desired power .80, the minimum sample size per group is 17. The sample sizes were thus more than adequate for detecting effects of this magnitude. Actual SD was 33 points, indicating estimated SD was conservative.

Despite the small sample sizes, 95 % confidence interval ranges were small enough to reach strong inferences for the first questionnaire due to the large effect, as predicted by pretesting. Because of an 18 % dropout rate in the second questionnaire and use of 20 statements in the first questionnaire versus 13 in the second, confidence intervals widened in the second questionnaire.

5.3. *Vote question results*

In Fig. 7, correct answers are 9 for deceptive and 1 for non-deceptive statements, under the idealized conditions stated in the vote question. Other answers indicate errors in LTQ or AAQ reasoning. Even if these errors are seemingly small and insignificant, they can have large cumulative effects when repeated exposure to persuasive stimuli occurs over time (Koch & Arendt, 2017), a cognitive flaw caused by the processing fluency effect (Linden, 2022, italics added):

> The primary cognitive mechanism responsible for the fact that people are more likely to think that repeated claims are true is known as *processing fluency*: the more a claim is repeated, the more familiar it becomes and the easier it is to process. In other words, the brain uses fluency as a signal for truth.

The study was single exposure, so we make no claim it tests cumulative effects. We rely on research by others that shows that small shifts in beliefs can accumulate into large shifts over time, due to repeated exposure to media messages.

In democratic systems, once a person knows a politician has lied and knows the Penalize the Deceiver rule, answers 6, 7, and 8 only partially penalize the deceiver and are an error in AAQ reasoning, as defined in the training. What is needed is *full* penalization, by applying the Penalize the Deceiver rule. This normative rule helps a person avoid cumulative media effects.

The error in AAQ reasoning occurs because of the tendency of *moral decoupling* (Bhattacharjee et al., 2012), where: "judgments of performance are separated from judgments of morality. By separating these judgments, moral decoupling allows consumers to support a transgressor's performance while simultaneously condemning their transgressions." Moral decoupling is a rationalization heuristic that allows people "to support public figures that act immorally." A voter choosing answer 8, for example, is saying "I know the politician lied and condemn them for that, but I still partially support them." Due to cumulative effects, repeated exposures can cause weak beliefs (such as partial support) to grow into strong beliefs.

Vote question answers show interesting behavior:

5.3.1. Group 1. Trained on neutral topic (control group)

While the effect surely varies across political units and study samples, we hypothesize that the first row approximates how voters in most nations behave today.

In chart A there's more support than opposition in response to a deceptive political claim. This has not gone unnoticed by politicians willing to engage in deception.

A normal distribution curve centered on the midpoint was added to chart A to illustrate how closely the data approximates, but does not perfectly fit, a normal distribution. Low DTQ has caused answers to shift slightly left. The near-normal shape of the data indicates a person's level of political truth literacy is largely due to random factors rather than the formal education seen in charts C and especially E. This may be inferred from the way a normal distribution results from a sufficiently large sample of random independent variable values. Chart A measures the informal assimilation discussed earlier in section 4.8, which we hypothesize to be largely random.

Chart B shows untrained people lean toward supporting truth-telling politicians who tell the truth, but seem shy about supporting them

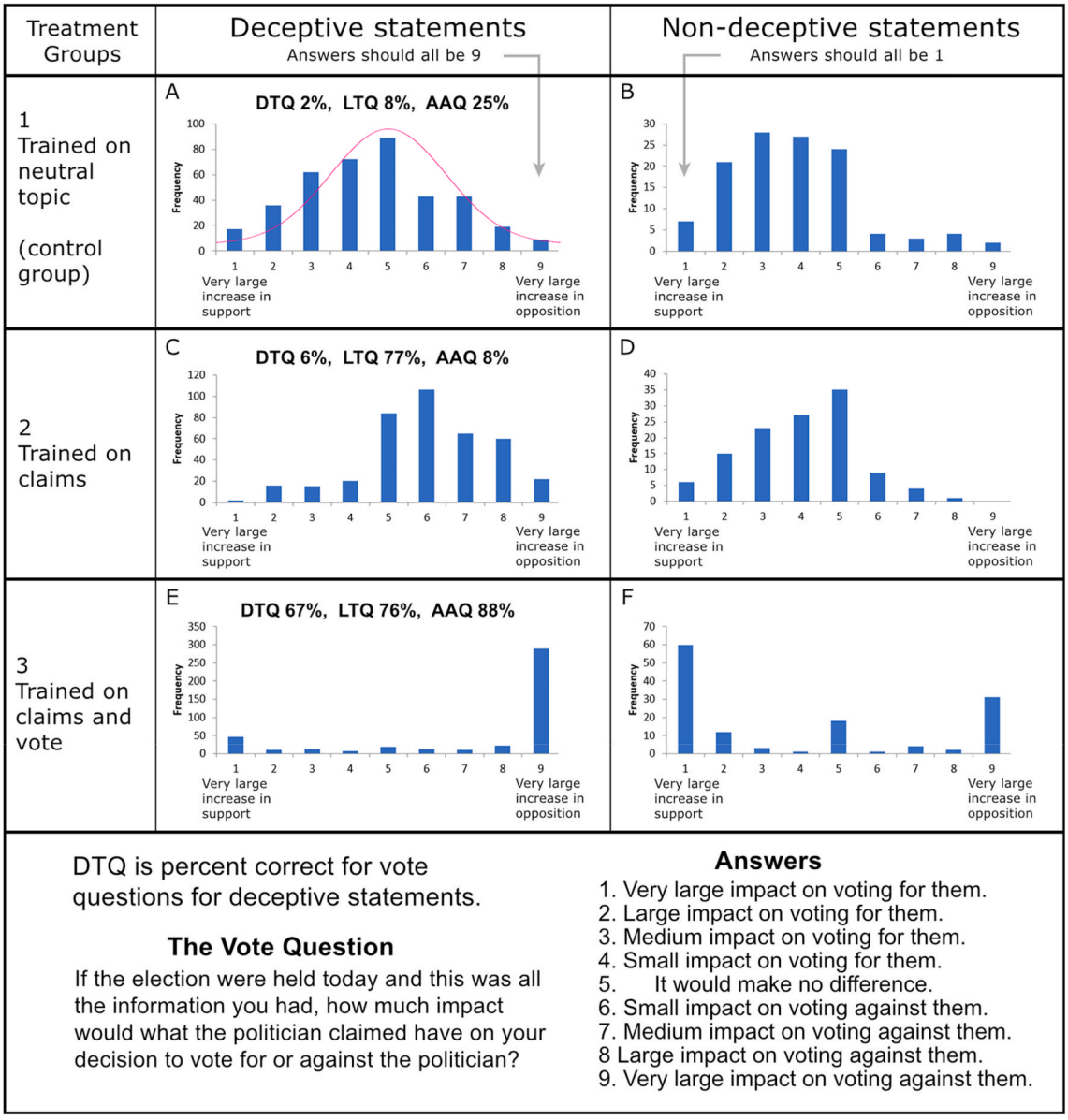

Fig. 7. Distributions of the vote question answers for the first questionnaire. The question measures DTQ for deceptive statements. Values for the equation DTQ = LTQ x AAQ are shown.

strongly with answer 1. We hypothesize this is mostly because they received no vote training.

5.3.2. Group 2. Trained on claims

As discussed in section 5.2, the second row contains what to us is astonishing counterintuitive data. Citizens trained on how to determine the truth of claims but not trained in how to vote correctly, intuitively lean in the correct direction on vote answers. *But very few choose the correct answers of 9 in chart C and 1 in chart D.* A surprising percentage (22 % and 29 %) choose answer 5, "It would make no difference." That's like saying "It doesn't matter to me at all if a politician tells the truth or not." But yet it must, if democracy is to thrive.

Similar observations apply to other incorrect answers. Choosing 4 and 6 is like saying "It barely matters to me whether a politician tells the truth or not." Deviations from correct answers are why the vote training in group 3 is required.

Correct answer preference is worse in chart D than chart B. Claim training *reduced* ability to vote correctly. This is puzzling behavior we cannot explain and identifies an interesting phenomenon for further research.

5.3.3. Group 3. Trained on claims and vote

The third row, if we could get enough voters to choose the correct answers when faced with real-world decisions, would resolve the root cause by raising political truth literacy components from low to medium. The simulation model shows levels of 50 % for LTQ and AAQ are sufficient to solve the problem. These levels appear feasible. To accelerate solution, training would go first to those most likely to benefit from it: swing voters and the young. The rest are committed to a party and rarely change (Jones et al., 2020).

The training needs improvement to reduce confusion of some kind, indicated by the answer 1 spike in chart E and the answer 2, 5, and 9 spikes in chart F. These should all be near zero.

6. Strengths and limitations

6.1. *Strengths of RCA*

The main benefits of an RCA-based process using something like social force diagrams are:

6.1.1. The method allows avoidance of the Superficial Solutions Trap.

Fig. 2 shows how common solutions to the vulnerability of democratic systems to political disinformation problem fell into the trap. Since social force diagrams are designed to apply to members of the class of difficult large-scale social problems, the tool can help social scientists escape the trap in members of this class, as well as less difficult social problems. However, reaching method maturity will take time.

We make this claim because the same pattern has occurred in industry. The pattern is that RCA is generic and historically has required process wrappers for each distinct problem domain. As RCA slowly diffused from where it started (auto manufacture in Japan in the 1950s) to all countries and all large-scale industries, new process wrappers were developed as needed.

6.1.2. The method allows construction of a comprehensive, sufficiently complete causal theory of a difficult large-scale social problem.

The theory explains the four social forces that characterize this class of problems and the essential causal structure (ECS) involved.

The more the ECS of a problem is known, the more likely the problem can be solved and the better it will be solved. This principle is embodied in the terms black, gray, and glass box models. In a *black box model*, "the computations are hidden and relationships between the variables of the system can only be inferred" (Blumschein et al., 2019, p. 9), such as correlational statistical models, where only the independent and dependent variables are known. In a *gray box model,* ECS is partially revealed, such as statistical models supplemented by field observations. In a *glass box model,* "which overtly displays all [relevant] mechanisms and functions of the system being modeled," ECS is sufficiently revealed, such as by use of social force diagrams or other forms of RCA. Applying critical realism, Pawson and Tilley (1997) argue that for complex social program evaluation, correlation models are insufficient, because causal mechanisms acting in contexts cause outcomes. Understanding those mechanisms and their contexts requires glass box modeling.

6.1.3. The method offers a potential starting point for the policy formulation step of the government policy process for difficult large-scale social problems.

The literature has repeatedly reported deficiencies in this step. Howlett (2007) frames this as the problem of low "policy analytical capacity [which contributes] to failure to effectively deal with many complex contemporary policy challenges." Analytical capacity is one of the five core elements of "the problem-solving capacity of the modern state," the others being delivery, coordination, regulatory, and management capacity(Lodge and Wegrich, 2014).

The method can potentially increase analytical capacity, as summarized in strengths 1 and 2. As described in the Method of Research supplementation material on *The five steps of applying social force diagrams*, the fifth step develops a small group of solution candidates that have a high probability of success. These become solution options.

6.2. Limitations of the analysis and experimental study

The chief limitation is that this is a combined feasibility and proof-of-concept project and is therefore greatly simplified, which imposes many constraints. A *feasibility study* is a preliminary investigation to determine if a research approach is conceptually possible (practical) and roughly how it could be accomplished. It focuses on process (including intervention design) rather than outcomes (Bowen et al., 2009; Eldridge et al., 2016). In this project the process is the analysis method and experiment and training questionnaire design. A *proof-of-concept study* takes the next step by empirically assessing if a process is potentially capable of producing intended outcomes under limited conditions. It focuses on outcomes (Craig et al., 2019; Eldridge et al., 2016). In this project the outcomes are the theory produced by the analysis method and a small-N controlled experiment to test key theoretical predictions. Both study types ask "Is this possible?" and serve to determine whether larger-scale, more-complete research is warranted.

Accordingly, the social force diagram and simulation model simplifies by omitting factors like multiple root causes, lie amplification, bots, AI micro-targeting, foreign state propaganda and intelligence operations, cyclic behavior, hereditary behavior, more than two groups in the left-right political spectrum, simulation model calibration (a large task in itself), much more detail in the diagram, etc.

Simplifications like these are normal for feasibility and proof-of-concept research, and fall within the methodology used. RCA operates on the strategy of finding a problem's essential causal structure and ignoring everything else. This can lead to surprisingly simple but useful causal structures, which in our case are the social force diagram and simulation model. These are surprisingly simple when the complexity of the problem is considered, and allowed high leverage point identification and creation of the key hypotheses needing testing, H1 and H2.

The Truth Literacy Training study simplifies by using a small sample size, a partially representative sample of a single country using an online panel, narrow training on only a few fallacies, a single follow up study, and use of text statements instead of other forms of persuasion like videos, images, and social media. Simplifications like these are typical for feasibility and proof-of-concept research. The study needs replication, larger more representative samples, broader training, deeper investigation of falloff effects with longitudinal questionnaires, testing in multiple nations, training and testing on real-world political statements and other forms of persuasion like videos, images, and social media, and real-world testing.

All these limitations are opportunities for further research.

7. Conclusions

Our main conclusion is that the analysis presented here offers a starting-point comprehensive theory of the vulnerability of democracy to political disinformation problem that can productively guide further research. The theory consists of the causal structures in Figs. 2 and 3, the cognitive model of DTQ = LTQ x AAQ, and the four explanations below. These are based on the "four requirements for a comprehensive theory of a difficult large-scale social problem" as discussed earlier.

1. **Why past solutions failed** (because S < R) – Due to lack of an appropriate analytical method, the fundamental layer of the problem was hidden by complexity. This caused problem solvers to be intuitively attracted to pushing on low leverage points with superficial solutions. The result was reliance on the superficial solutions shown in Fig. 2.
2. **Why the problem occurs** (because R is unresolved) – The current system cannot achieve common good goals because of the unresolved main root cause of *low political truth literacy*, as measured by the Truth Literacy Training study.
3. **Why fundamental solutions can be expected to succeed** (because F > R) – The Truth Literacy Training study provides preliminary proof that the root cause can be resolved in a practical manner, though much further research is required. *Solution strategy must center on resolving the main root cause by raising the components of DTQ (LTQ and AAQ) from low to at least medium.*
4. **Why the mode change will be relatively permanent** (because new R contains self-sustaining feedback loops) – Resolving the root cause changes feedback loop dominance and how politicians compete for votes, causing the system to shift into the new mode where *politicians constructively compete to see who can tell the most helpful truths about how to optimize the long-term common good.* This should be relatively permanent, because based on this analysis and future similar research, those who believe in democracy now know that keeping political truth literacy *above a critical minimum* is required for a healthy democracy and will strive to keep it there. How this rule of system behavior can be institutionalized is an area for further research.

That DTQ must be kept above a critical minimum is what Lynch (2025, p. 18) implies with "a [correctly] informed citizen is essential for democracy to flourish." Dahl's (2015, p. 37) third criterion for a democratic process, "enlightened understanding," echoes that requirement, as does Lewandowsky's (2024) claim that "Democracy also requires reliable shared knowledge for meaningful debate and to ensure normatively good policy outcomes." Low DTQ prevents correct understanding, enlightened action, and reliable shared knowledge.

A second conclusion is multiple solution elements will be required for effective full root cause resolution, for these reasons and more:

1. Many news sources are untrustworthy and are getting better at appearing just as reliable as trustworthy sources.
2. Many claims are too complex to easily evaluate with the Personal Truth Test, such as "This budget will work."
3. Many citizens will feel too busy or distracted to take the time to consciously apply the training, will lose motivation to prevent being fooled, will reach different levels of competency after training, or may not recognize their own cognitive weaknesses.
4. There will be intense change resistance and ingenious adaptation from those working for the Race to the Bottom. They cannot afford to let the majority learn how to tell truth from deception. For example, use of AI will trigger an arms race between people's growing DTQ skills and the deceivers' AI-designed disinformation that's harder to detect with LTQ, confuses the AAQ response, and is so well micro-targeted that a person's defenses are weakened.
5. The mushrooming amount of fake news and disinformation in social media exists because it maximizes profits and drives the business model of "digital capitalism." Curtailing this requires government regulation (Buckingham, 2019).

These factors point to the need for research on more than just Truth Literacy Training, such as News Source Truth Ratings, Politician Truth Ratings, Continuing Truth Literacy Training in Journalism by making the deception mechanism a central part of the story, the legal right to Freedom from Falsehood from public servants the public must be able to trust (such as politicians and information sources the public is dependent on), a searchable online database of AI generated fact-checks that can also be generated on request, a phone app to run the Personal Truth Test on a claim, and so on. AI will soon be able to do most of the work for some of these solution elements. Furthermore, AI-powered Truth Literacy Training using an educational tool like Khan Academy's Khanmigo *is now feasible and could be rapidly implemented at low cost* (Tyrangiel, 2024).

CRediT authorship contribution statement

Jack Harich: Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Conceptualization. **Montserrat Koloffon Rosas:** Writing – review & editing, Investigation, Conceptualization.

Ethical approval and informed consent

The questionnaire study received ethical approval from an appropriate review group at Thwink.org. Participants were anonymized by Prolific.com. All participants read this privacy notice at the beginning of the questionnaire: "This survey is run by Thwink.org. All data will be held strictly confidential and is used only for research. No personally identifiable data is collected." While this was not a medical study, ethical principles congruent with the WMA Declaration of Helsinki Ethical Principles for Medical Research Involving Human Participants were followed.

Ethical approval and informed consent

The questionnaire study received ethical approval from an appropriate review group at Thwink.org. Participants were anonymized by Prolific.com. All participants gave their informed consent via the online form.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work the authors used ChatGPT and Claude for the final language edit, for typos, spelling, grammar, and to spot wording problems. After running checks, the authors reviewed results and decided what to do, and take full responsibility for the content of the published article. No AI was used in the writing of this article, the Figures, or the supplementaty materials.

Declaration of competing interest

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

Acknowledgements

Special thanks to George Turner, Shayne Gary, Markus Schwaninger, and especially Philip Bangerter for assistance in the long gestation of this complex research. We also thank the journal editor and three reviewers for their painstaking review of this very complex manuscript.

Appendix A. Supplementary material

This includes model description, method of analysis, the simulation model, study data, and questionnarie description. These files can be found online at https://doi.org/10.1016/j.ssaho.2026.102469.

Data availability

See Supplementary materials for the study data.

References

Baines, D., Brewer, S., & Kay, A. (2020). Political, process and programme failures in the brexit fiasco: Exploring the role of policy deception. *Journal of European Public Policy, 27*(5), 742–760. https://doi.org/10.1080/13501763.2020.1722203

Banjanovic, E., & Osborne, J. (2016). Confidence intervals for effect sizes: Applying bootstrap resampling. *Practical assessment, research & evaluation*.

Bateman, J., & Jackson, D. (2024). *Countering disinformation effectively: An evidence-based policy guide*. Carnegie Endowment for International Peace.

Bell, M., Whitehead, A., & Julious, S. (2018). Guidance for using pilot studies to inform the design of intervention trials with continuous outcomes. *Clinical Epidemiology, 10*, 153–157.

Benkler, Y., Faris, R., & Roberts, H. (2018). *Network propaganda: Manipulation, disinformation, and radicalization in American politics*. Oxford University Press.

Bhattacharjee, A., Berman, J., & Reed, A. (2012). Tip of the hat, wag of the finger: How moral decoupling enables consumers to admire and support immoral actors. *Journal of Consumer Research, 39*, 1167–1184.

Bhattachewrjee, A. (2012). *Social science research: Principles, methods, and practices (second)*. Scholar Commons.

Blumschein, P., Hung, W., Jonassen, D., & Strobel, J. (2019). *Model-based approaches to learning: Using systems models and simulations to improve understanding and problem solving in complex domains*. Sense Publishers.

Bowen, D., Kreuter, M., Spring, B., Cofta-Woepel, L., Linnan, L., & Weiner, D. (2009). How we design feasibility studies. *American Journal of Preventative Medicine, 36*, 452–457.

Brodie, R. (1996). *Virus of the mind: The new science of the meme*. Hay House.

Buckingham, D. (2019). Teaching media in a 'Post-Truth' age: Fake news, media bias and the challenge for media literacy education. *Cultura Y Educación, 31*(2), 1–19.

Byers, J., Mitzenmacher, M., & Zervas, G. (2012). The groupon effect on yelp ratings: A root cause analysis. *Proceedings of the 13th ACM conference on electronic commerce*. https://arxiv.org/pdf/1202.2369.

Civicus, S. (2023). Finland tilts to the far right. *CIVICUS Lens*. https://lens.civicus.org/finland-tilts-to-the-far-right/.

Conason, J. (2004). *Big lies: The right-wing propaganda machine*. St: Martins Press.
Conley, R., Hall, K., & Scott, L. (2024). *Racial disparity root cause analysis for the department of the air force. RAND*. https://www.rand.org/pubs/research_reports/RRA1574-1.html.
Coskun, R., Akande, A., & Renger, R. (2012). Using root cause analysis for evaluating program improvement. *Evaluation Journal of Austrtalasia, 12*, 4–14.
Craig, P., Dieppe, P., Macintyre, S., Michie, S., Nazareth, I., & Perricrew, M. (2019). Developing and evaluating complex interventions: The new medical research council guidance. *BMJ, 337*, a1655.
Dahl, R. (2015). *On democracy*. Yale University Press.
Dawkins, R. (1976). *The selfish gene*. Oxford University Press.
Dillard, J., & Shen, L. (2013). *The SAGE handbook of persuasion: Developments in theory and practice*. SAGE.
DRL Staff. (2023). Duke Reporter's lab: Misinformation spreads, but fact-checking has leveled off. https://sanford.duke.edu/blog-post/duke-reporters-lab-misinformation-spreads-fact-checking-has-leveled/.
Dumitru, E.-A., Ivan, L., & Loos, E. (2022). A generational approach to fight fake news: In search of effective media literacy training and interventions. *International Conference on Human-Computer Interaction*, 291–310. https://doi.org/10.1007/978-3-031-05581-2_22.
Dunlap, R., & Jacques, P. (2013). Climate change denial books and conservative think tanks: Exploring the connection. *American Behavioral Scientist, 57*, 699–731.
Dunlap, R., & McCright, A. (2011). Organized climate change denial. In *The Oxford handbook of climate change and society* (pp. 144–160). Oxford University Press.
Eldridge, S., Lancaster, G., Campbell, M., Thabane, L., Hopewell, S., Coleman, C., & Bond, C. (2016). Defining feasibility and pilot studies in preparation for randomised controlled trials. *PLoS One, 11*, Article e0150205.
Erlanger, S., & Anderson, C. (2022). Rise of far-right party in Sweden was both expected and shocking. *The New York Times*. https://www.nytimes.com/2022/09/15/world/europe/sweden-election-far-right.html.
Etzerpdt, S., & Kohgshoj, K. (2022). The implosion of radical right populism and the path forward for social democracy: Evidence from the 2019 Danish national election. *Scandinavian Political Studies, 45*, 279–300.
Freelon, D., Marwick, A., & Kreiss, D. (2020). False equivalencies: Online activism from left to right. *Science, 369*(6508), 1197–1201.
Gidron, N., & Hall, P. (2017). The politics of social status: Economic and cultural roots of the populist right. *British Journal of Sociology, 68*, 57–84.
Graves, L. (2016). *Deciding what's true: The rise of political fact-checking in American journalism*. Columbia University Press.
Guess, A., & Nyhan, B. (2018). *Selective exposure to misinformation: Evidence from the consumption of fake news during the 2016 U.S. presidential campaign* (Vol. 49). European Research Council.
Hameleers, M. (2022). Separating truth from lies: Comparing the effects of news media literacy interventions and fact-checkers in response to political misinformation in the US and Netherlands. *Information, Communication & Society, 25*, 110–126.
Hesk, J. (2000). *Deception and democracy in classical Athens*. Cambridge University Press.
Hobbs, R. (2010). Digital and media literacy: A plan of action. *The aspen institute*.
Hobbs, R. (2020). *Mind over media: Propaganda education for a digital age*. Norton.
Howlett, M. (2007). *Policy analytical capacity as a source of policy failure. Workshop on policy failure*. Canadian Political Science Association Annual Meeting.
Jackson, B., & Jamieson, K. H. (2007). *Unspun: Finding facts in world of disinformation*. Random House.
Jones, B., Doherty, C., Daniller, A., Dunn, A., & Green, T. (2020). *Voters rarely switch parties, but recent shifts further educational and racial divergence*. PEW Research Center. https://www.pewresearch.org/politics/2020/08/04/voters-rarely-switch-parties-but-recent-shifts-further-educational-racial-divergence/.
Kettle, M. (2024). The shocking US election result will create a new world order – And launch a fresh wave of trump wannabes. *The Guardian*. https://www.theguardian.com/commentisfree/2024/nov/06/us-election-donald-trump-new-world-order.
Koch, T., & Arendt, F. (2017). Media effects: Cumulation and duration. In P. Rössler, C. A. Hoffner, & L. Zoonen (Eds.), *The international encyclopedia of media effects* (1st ed., pp. 1–11). Wiley. https://doi.org/10.1002/9781118783764.wbieme0217.
Kurbjuweit, D. (2024). Donald trump and the new world order: The end of the west. *Der Spiegel*. https://www.spiegel.de/international/world/donald-trump-and-the-new-world-order-the-end-of-the-west-a-b71fa1bd-6147-47a4-8738-5c17eff44a55.
Larrabee, H. (1952). *Bentham's handbook of political fallacies*. John Hopkins Press.
Lessenski, M. (2022). *How it started, how it is going: Media literacy index 2022*. European Policies Initiative. https://osis.bg/wp-content/uploads/2022/10/HowItStarted_MediaLiteracyIndex2022_ENG_.pdf.
Lewandowsky, S. (2024). Truth and democracy in an era of misinformation. *Science, 386*, eads5695.
Liagusha, A., & Iarovyi, D. (2025). Memes, freedom, and resilience to information disorders: Information warfare between democracies and autocracies. *Social Sciences & Humanities Open, 11*, 101247.
Liker, J. (2004). *The toyota way: 14 management principles from the world's greatest manufacturer*. McGraw Hill.
Linden, S. (2022). Misinformation: Susceptibility, spread, and interventions to immunize the public. *Nature Medicine, 28*, 460–467.
Lodge, M., & Taber, C. (2013). *The rationalizing voter*. Cambridge University Press.
Lodge, M., & Wegrich, K. (2014). *The problem-solving capacity of the modern state: Governance challenges and administrative capacities*. Oxford University Press.
Luhrmann, A., & Lindberg, S. (2019). A third wave of autocratization is here: What is new about it? *Democratization, 26*(7), 1095–1113. https://doi.org/10.1080/13510347.2019.1582029
Luhrmann, A., Medzihorsky, J., & Lindberg, S. I. (2023). Walking the talk: How to identify anti-pluralist parties. *Party Politics, 30*, 420–434.
Lynch, M. (2025). *On truth in politics: Why democracy demands it*. Princeton University Press.
McDougall, J., Zezulkova, M., van Driel, B., & Sternadel, D. (2018). *Teaching media literacy in Europe: Evidence of effective school practices in primary and secondary education (NESET II report)*. NESET.
Moore, T. (2013). Critical thinking: Seven definitions in search of a concept. *Studies in Higher. Education, 4*, 505–522.
OECD. (2023). *PISA 2022 results volume 1: The state of learning and equity in education*. OECD Publishing. https://doi.org/10.1787/53f23881-en
Ohno, T. (1988). *Toyota production system: Beyond large-scale production*. Productivity Press.
Pawson, R., & Tilley, N. (1997). *Realistic evaluation*. SAGE.
Percarpio, K., Watts, V., & Weeks, W. (2008). The effectiveness of root cause analysis: What does the literature tell Us? *Joint Commission Journal on Quality and Patient Safety, 34*, 391–398.
Potter, W. J. (2022). *Analysis of definitions of media literacy*. Journal of Media Literacy Education.
Powell, L. (2023). *Fake news and critical literacy: Final report of the commission on fake news and the teaching of critical literacy in schools*. National Literacy Trust. https://nlt.cdn.ngo/media/documents/Fake_news_and_critical_literacy_-_final_report.pdf.
Prooijen, J.-W., Rodrigues, T., Bunzel, C., Georgescu, O., Komaromy, D., & Krouwel, A. (2022). Populist gullibility: Conspiracy theories, news credibility, bullshit receptivity, and paranormal belief. *Political Psychology, 43*, 1061–1079.
Puddington, A. (2017). *Breaking Down democracy: Goals, strategies, and methods of modern authoritarians*. FreedomHouse.org. https://freedomhouse.org/sites/default/files/2020-02/June2017_FH_Report_Breaking_Down_Democracy.pdf.
Rooduijn, M., & Burgoon, B. (2018). The paradox of well-being: Do unfavorable socioeconomic and sociocultural contexts deepen or dampen radical left and right voting among the less well-off? *Comparative Political Studies, 51*, 1720–1753.
Roozenbeek, J., Culloty, E., & Suiter, J. (2023). Countering misinformation: Evidence, knowledge gaps, and implications of current interventions. *European Psychologist, 28*, 89–205.
Sagarin, B. J., Cialdini, R. B., Rice, W. E., & Serna, S. B. (2002). Dispelling the illusion of invulnerability: The motivations and mechanisms of resistance to persuasion. *Journal of Personality and Social Psychology, 83*(3), 526–541. https://doi.org/10.1037/0022-3514.83.3.526
Scoones, I., Edelman, M., Borras, S., Forero, L., Hall, R., Wolford, W., & White, B. (2020). *Authoritarian populism and the rural world*. Palgrave Macmillan.
Singleton, R., & Straits, B. (2005). *Approaches to social research, fourth edition*. Oxford University Press.
Sterman, J. (2000). *Business dynamics: Systems thinking and modeling for a complex world*. Irwin McGraw-Hill.
Tyrangiel, J. (2024). An 'education legend' has created an AI that will change your mind about AI. *Washington Post*. https://www.washingtonpost.com/opinions/2024/02/22/artificial-intelligence-sal-khan/.
Wagner, T. (2014). Using root cause analysis in public policy pedagogy. *Journal of Public Affairs Education*.
Weinstein, J., Geller, A., & Negussie, Y. (2017). *Communities in action: Pathways to health equity*. The National Academies Press.
Wikipedians. (2024). Danish people's party. In *Wikipedia*. https://en.wikipedia.org/wiki/Danish_People%27s_Party#Parliament.
Wood, A., Stedman-Edwards, P., & Mang, J. (2000). *The root causes of biodiversity loss*. Earthscan Publications.