Truth Literacy Training: The core solution element for resolving the main root cause of democratic backsliding

Jack Harich¹, Montserrat Koloffon Rosas²
¹Thwink.og, Atlanta, USA. jack@thwink.org. Corresponding author.
²Vrije Universiteit Amsterdam, Amsterdam, Netherlands.

Abstract

A precipitous backward slide from democracy to authoritarianism/autocracy is underway. Root cause analysis and a system dynamics simulation model were used to analyze why this occurred and how the backward slide can be reversed. The main root cause was found to be low political truth literacy. As long as this is low, democracy cannot function as intended because citizens are too easily deceived into voting against their own best interests. The paper reports on an empirical study of what appears to be the core solution element for raising political truth literacy: Truth Literacy Training. Results indicate average political truth literacy is currently low (the root cause exists) and can be raised to high with a relatively small amount of carefully designed training (the root cause can be resolved). A collection of solution elements is required for complete root cause resolution.

Introduction

After centuries of democracy’s long fitful rise as an alternative to autocracy, capped by collapse of the Soviet Union in 1991, Fukuyama¹ famously declared “the end of history” had arrived: Liberal democracy had proven itself superior to all other forms of government and would eventually become the universal norm, based on the widespread assumption that high economic development requires the efficient mechanisms of liberal democracy.² However, this optimistic prediction was shattered by arrival of a steep slide toward autocratization beginning in the late 1990s, as so clearly illustrated in Figure 1.

Why did this happen? Why is democracy backsliding to authoritarianism?

The topic of “why democracies break down” has drawn “huge amounts of attention” from scholars.³ However, despite this effort “we lack theories to explain backsliding, though we have long engaged in a perhaps interminable debate about the causes of democratic transitions, democratic breakdowns, authoritarian resilience, and democratic consolidation”.⁴ Specifically, “there is a lack of work on the question of under which circumstances might instability arise”.⁵ The literature offers no satisfactory comprehensive explanation of democratic backsliding and thus no clear path to analytical solution.
This gap was addressed in a previous paper (Add reference after analysis paper accepted), hereafter called “the analysis paper.” Using root cause analysis and system dynamics simulation modeling, the analysis paper presented an in-depth analysis of the causal structure of the problem, highlighted by the problem’s main root cause and high leverage point. The key assumption of the analysis, that low political truth literacy is the main root cause of the democratic backsliding problem, was empirically tested by the Truth Literacy Training (TLT) study. Political truth literacy is the ability to tell truth from falsehood in political statements meant to influence voters and to make logically correct voting decisions based on that information. The analysis and study provide a complete theory, based on identification of the problem’s causal structure and preliminary empirical evidence, that explains why backsliding occurs, why past solutions have failed, and how the problem can be solved in a practical manner.

The analysis paper presented the complete analysis and a short summary of the TLT study. This paper presents a summary of the analysis and the full study, with the qualification that TLT is a partial solution element for resolving the main root cause of democratic backsliding. It is not a sufficient solution. A collection of solution elements will be required to fully resolve the main root cause, as well as any additional root causes found in subsequent research.

The analysis and solution should not be interpreted as the analysis or the solution, but as an example of what is possible. The deeper message of the two papers is that the analysis and solution are an example of how difficult large-scale social problems like backsliding can be analysed and solved in an engineering-like manner by use of appropriate tools. To clarify the message this paper elaborates on the analysis method.

The remainder of this paper begins with a summary of analysis, followed by full description of the TLT study, additional solution elements, and conclusions.
Summary of analysis

Method

The standard method for solving difficult business problems is some version of root cause analysis, since all causal problems arise from their root causes. 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 bigger problems. Resolved means the problem will probably not recur due to that root cause. Root cause analysis is the systematic practice of finding, resolving, and preventing recurrence of the root causes of causal problems.9

Surveying the business and academic literature, we found no root cause analysis-based method was available for difficult social problems so we were compelled to develop one, a common occurrence on novel classes of problems. The result was social force diagrams. Similar to cause-and-effect (fishbone) diagrams,10 social force diagrams use a standard fill-in-the-blanks template for diagramming the high-level causal structure of a difficult large-scale social problem in an efficient manner.

Figure 2 contains the social force diagram for the backsliding problem. The diagram is organized into two layers: (1) the superficial (symptomatic) layer of the problem, where intermediate causes are so easy to see they are routinely assumed to be root causes, and (2) the deeper fundamental layer, where by understanding the problem’s feedback loop structure its true root causes may be found. Social force diagrams are built by starting at problem symptoms and working backward with “WHY does this occur?” questions until the root causes are found. As this is done, why past superficial solutions have failed is diagrammed. This is important knowledge, as it indicates the intermediate causes are indeed intermediate rather than root causes. After the superficial layer of the problem is understood, the analyst follows the causal chain down into the fundamental layer to find the root causes and fundamental solutions.

Knowledge of the superficial layer and why past solutions failed is mandatory for solving difficult problems, because as Karl Popper11 explains:

We are always learning a whole host of things through falsification. We learn not only that a thing is wrong; we learn why it is wrong. Above all else, we gain a new and more sharply focused problem, and a new problem, as we already know, is the starting point for a new development in science.

After the superficial layer is built a new problem that could not be seen before comes into sharp focus: What is the feedback loop structure of the fundamental layer? Where in that structure is the root cause of the lowest intermediate cause in the superficial layer? What is the high leverage point for resolving the root cause? What practical solutions can push on the high leverage point?

Social force diagrams center on understanding three key forces. Superficial solutions (force S) fail because force S is always less than root cause forces (force R), indicated on Figure 2 by S<R. By contrast, fundamental solutions (force F) can succeed because if the solutions are properly designed, force F can exceed force R, indicated by
F>R. Once all three forces are understood and key assumptions have been measured or tested, the analyst has a sufficiently complete theory of problem behaviour: Why the problem occurs (force R is unresolved), why present solutions are failing (S<R), and why fundamental solutions can be expected to succeed (F>R).

**Social force diagram of the democratic backsliding problem**

![Social force diagram of the democratic backsliding problem](image)

Figure 2. Social force diagram of the analysis. A simpler but less accurate term for the main root cause is *low political truth literacy*. Use of feedback loops in social force diagrams is optional. The superficial layer contains one or more intermediate causes. Understanding the causal structure of the fundamental layer requires a feedback loop model. Some problems require multiple diagrams, since they contain multiple subproblems (defined by multiple symptoms) and thus multiple root causes.

The diagram was developed in this manner: After summarizing problem symptoms as *backsliding from democracy to authoritarianism*, we asked: WHY do these symptoms occur? Svolik\(^1\) reports that 197 democratic backslides occurred from 1973 to 2018. Of these, 46 were military coups and 88 were executive takeovers via election, with takeovers averaging about 80% of all backslides after the end of the Cold War in 1991. In an examination of *How Democracies Die*, Levitsky and Ziblatt\(^2\) summarize this change: “Democratic backsliding today begins at the ballot box.” Thus, the main first intermediate cause of backsliding is *election of politicians not working for the democratic common good*.

The solution used now is the same solution used before to hasten the spread of democracy before the backslide began. The solution attempts to *promote and prove the superiority of democracy over authoritarianism*. This is a form of *more of the truth*. The solution no longer works because since about 2000, authoritarian state capacity (government effectiveness, regulatory quality, rule of law, and control of corruption)
has improved so much worldwide, especially in China, that there is no longer proof democracy is superior.\textsuperscript{14}

Next, we asked: WHY does election of politicians not working for the common good occur? The answer is mainly because of successful political deception. Other factors (discussed in the next section) are much less effective. Deception is the main technique used to convince an electorate majority to act against their own best interests.

The evidence shows that all authoritarians depend heavily on deceptively provoking a wide range of false beliefs and emotions, especially fear. McCarthy\textsuperscript{15} describes how this involves “bellicose rhetoric,” false promises of forceful solutions to complex long-term problems, populist and racist appeals, identification of false internal enemies, and more.

Finally, we asked: WHY does successful political deception occur? Answering this required construction of a system dynamics simulation model (Figure 3). The answer was because of the inherent advantage of the Race to the Bottom, summarized as low political truth literacy. Since it appears this can be resolved by pushing on the high leverage point of raise political truth literacy from low to high, this is the main root cause.

\textit{System dynamics model}

Model construction was driven by understanding why successful political deception occurs. The model captures the essence of the left-right political spectrum, consequential because “global politics is first and foremost a debate between the left and the right. ... The left-right dichotomy occupies a special place, as the most enduring, universal, and encompassing of all political strategies”.\textsuperscript{16}
Figure 3. Feedback loop structure of the backsliding problem. This is a simplified version of the full system dynamics model. The backbone of the structure is the two opposing feedback loops dueling for the same Uncommitted Supporters. Race to the Bottom politicians use deception to gain supporters, while Race to the Top politicians use the truth. Solid arrows are a direct relationship. Dashed arrows are an inverse relationship. R and B signify reinforcing and balancing loops.

The Race to the Bottom feedback loop (the right) represents powerful special interests pursuing their own narrow self-interest goals, such as the rich, authoritarians, man-agers of large for-profit corporations, and elite ruling groups of many kinds, e.g., the ruling class. All are a small percentage of the electorate. In a democracy, the main ways a minority can persuade a majority to vote for them are by force, threats/intimidation, rigged elections, voter suppression, favouritism, or deception. Force, threats, and rigged elections are illegal. Voter suppression is mostly illegal. Favouritism is inefficient, as even the rich lack the resources to bribe millions of voters. This leaves deception as the main preferred strategy and explains why deception is so common in right-wing politics. Jeremy Bentham, the father of utilitarianism, reached the same conclusion in 1824:

…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.17

The Race to the Top loop (the left) represents those seeking to cooperate in optimizing the common good of all. Politicians appealing to the left use a strategy of the truth about how they can achieve that goal.
The model uses Dawkins’ concept of memes. A meme is copied information capable of affecting behaviour, such as a fact or an opinion. In the model a meme is a statement that is true or false.

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 by saying that $2 + 2 = 5$, or $7$, or even $27$, but the size of the truth is always $1$. It 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 of backsliding: the inherent advantage of the Race to the Bottom, represented on the model by undetected false memes. The inherent advantage exists because the opposing loop, the Race to the Top, has no corresponding variable because there are no inflated true memes to detect. For simplicity, we usually say the main root cause is low political truth literacy.

Truth literacy is the ability to tell truth from deception, to be able to “read” the truth. The model and study use three important variables related to truth literacy. All range from zero to 100%:

1. LTQ (logical truth quotient) is the ability to logically tell if a deceptive 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 deceptive political claim.
3. DTQ (democratic truth quotient, aka political truth literacy) is the ability to take correct democratic system action (such as voting correctly) given a 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). Because of this process, the three variables are related by $DTQ = LTQ \times AAQ$, though DTQ is not in the model.

The single high leverage point in the social force diagram (DTQ) becomes two high leverage points in the model: LTQ and AAQ. As LTQ rises, so does detected false memes. That times AAQ equals actionable false memes. Actionable means “I need to take action because of all the lies I’ve detected.” As LTQ and/or AAQ rises, so does actionable false memes. As that rises two things happen: undetected false memes decrease, which reduces the power of the Race to the Bottom, and true memes increase, which strengthens the Race to the top.

The key analysis assumption was that the main root cause exists and can be resolved in a practical manner. To test this hypothesis, we performed:
The Truth Literacy Training (TLT) study

Solution element construction and study design

Using our own software and database, a form of online TLT was developed using a long questionnaire supplemented by training materials (Figure 4). An elaborate series of pretests was used to refine training and testing design. At one point 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. The authors found it was not enough to inoculate subjects by exposure to deceptive statements and explanation 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 authorities [a form of deception] 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 people that 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 Figure 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 (the first two are usually 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 deception, because we are so used to processing true statements from people we talk to, books we read, and so on.

This shocks people into realizing 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, 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.
Figure 4. Truth Literacy Training web page, group 3, claim and vote training. The subject has completed the training, done in the Getting Started and Review Section. They have just answered three questions concerning a statement about a Trade Agreement Treaty. The claim in the statement is bolded. All three answers are correct. The Personal Truth Test is shown on the right panel. Notes are the two vote training rules. Above The Personal Truth Test are summaries of the fallacies groups 2 and 3 were trained on. The subject scrolls to see all Reference Material. Using the left panel, subjects can navigate anywhere in the questionnaire to review their work. Checks indicate a completed item.

Using a Prolific online panel, the study was run on 93 US subjects randomly assigned to three groups. Demographics were age range 22 to 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 groups were:
**Group 1** (the control group) received training on the neutral topic of how democracy works.

**Group 2** received training on how to tell if a political claim (embedded in a political statement, such as the one in Figure 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 (Figure 5).

**Group 3** received the same training as group 2 plus training on how to vote correctly (given the perceived level of truth of a claim) by applying two rules: Reward the Truth Teller and Penalize the Deceiver (See Figure 4 Notes section for these rules.). Total time for group 3 averaged 87 minutes, of which about one hour was training. Group 3 training involves 37 questions.

There is a 5-minute break after training for all groups, necessary to avoid fatigue and loss of interest on such a long questionnaire. A follow up study was run later using different statements.

In the test section of the questionnaire (called Decision Making Section in Figure 4), non-hot statements were presented in random order. Figure 4 shows how each statement is 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 fictitious country of “Rutania” was used in statements to create interest and political realism without the 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.

In the study, LTQ was measured by the percent correct for the truth questions for deceptive statements. DTQ was measured by percent correct for vote questions for deceptive statements. AAQ is calculated from LTQ and DTQ.

Individual DTQ can theoretically never be higher than LTQ, since DTQ uses the results of LTQ as input. Study results support this prediction.

![The Strong Evidence Rule](image-url)

**Figure 5. The Strong Evidence Rule.**
General results

![Image of bar charts showing results of the Truth Literacy Training study.](image)

**Figure 6.** Results of the Truth Literacy Training study. Average scores and 95% confidence intervals for answers to deceptive statements are shown, with guessing levels, Cronbach’s alpha, and calculated AAQ. (AAQ = DTQ / LTQ) Treatment groups were:

1. – Trained on neutral topic (control group)
2. – Trained on claims
3. – Trained on claims and vote

Figure 6 summarizes study results for the three groups. Group 1 (the control group) corresponds to where the average political system is today. Political truth literacy (DTQ) is low. Group 2 claim training raises LTQ to high but as expected has little effect on DTQ, which remains still low. Group 3 claim and vote training raises both LTQ and DTQ to high.

The key data is LTQ and DTQ for groups 1 and 3. The large increases, from 8% to 76% and from 2% to 67%, 68-point and 65-point rises, suggest that TLT and other solution elements are capable of pushing on the high leverage point of *raise political truth literacy from low to high* successfully. Group 3 training averaged about one hour, indicating that TLT, such as in education systems and online training, will not require that much of a person’s time.

The follow up study 26 days later found LTQ and DTQ for group 3 had declined from 76% to 66% and 67% to 60%, 10-point and 7-point falls. After an average of 30 minutes 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 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 by regular exposure to political truth literacy topics in news media, such as a story centered on deceptive use of the cherry picking fallacy.

However, LTQ for group 1 was 22% and 20% for the pre-refresh and post-refresh training statements, versus 8% for the first questionnaire. This indicates spotting deception was substantially easier in the second questionnaire statements, and suggests there was more than the 10-point and 7-point declines noted above and that the refresh training may not have worked as well as the 75% and 70% indicated. The reason for this appears to be that the second questionnaire statements were developed after the first questionnaire was run. Without realizing it, we structured them slightly differently and frequently omitted stating how strongly supported the premises were. This caused the second set to be substantially easier than the first set, as it made fallacies easier to spot.

The problem is easily corrected. A more accurate measure of training persistence would require further statement testing/development and rerunning the study using balanced statements of equal difficulty in the first and second questionnaires. During this work the first and second questionnaires could be improved as needed.

**Hypotheses results**

In the first questionnaire, N=30, 30, and 33 for groups 1, 2, and 3. In the second questionnaire 26 days later, N=25, 27, and 24, an average dropout rate of 18%. A minimum sample size of 30 is required to assume results have a normal distribution and a reliable standard deviation can be calculated. Thus, the results of the first questionnaire are potentially reliable, while results for the second are slightly unreliable. It was our error to not use higher sample sizes for the first questionnaire to accommodate dropouts in the second questionnaire. Given the small sample sizes, we consider this a very preliminary study.

The study was designed to test six hypotheses. We reached these conclusions:

**H1.** LTQ and DTQ can be accurately measured.

Cronbach’s alpha is the standard measure of internal consistency, and hence reliability, of a set of Likert scale questions designed to measure a single construct (like LTQ or DTQ) in a single situation (such as groups 1, 2, and 3). The standard rules of thumb for alpha are: > .9 Excellent, > .8 Good, > .7 Acceptable, > .6 Questionable, > .5 Poor, and < .5 Unacceptable.

The trend in the first and second questionnaires is that the more the training, the higher the alpha. The more a person is trained to make a discrimination, like spotting deception and deciding how to vote, the less they must guess when choosing an answer. Less guessing increases internal consistency, which increases alpha. The relatively high alphas for group 3 in the first questionnaire, .82 and .92, indicate the scale itself is sound. LTQ and DTQ can therefore be accurately measured for group 3, the trained population.

But what of group 1, the untrained population, which has alphas of .38 and .44 in the first questionnaire? Increasing the number of questions in a survey of this type (where alpha is high enough to be promising and under some situations is high, like group 3) can be expected to increase alpha to an acceptable level, as long as the new
questions measure the same construct in a similar manner to the old questions. Thus, the results demonstrate that LTQ and DTQ can be accurately measured in general.

In the remaining hypotheses, only the 95% confidence intervals are considered.

**H2.** *LTQ and DTQ are currently low in the average voter.* This was well supported by the first questionnaire. Group 1 represents the average voter, who has never received the equivalent of TLT. Their LTQ and DTQ were low, 8% and 2%.

**H3.** *LTQ and DTQ can be raised to high via TLT.* This was also well supported by the first questionnaire. Because of training, LTQ rose from 8% to 77% and 76%. DTQ rose from 2% and 6% to 67%.

**H4.** *Training on LTQ alone is insufficient to raise DTQ to above the minimum DTQ for a healthy sustainable democracy.* This was well supported by the first questionnaire. The average vote score for those receiving claim training alone was 6%. This shot up to 67% for the group receiving claim and vote training.

The 6% 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 penalize deceptive politicians. Yet in a time when political deception is so rampant and the truth is so rare, why would anyone not want to strongly penalize deceivers? Why would anyone not want to strongly reward truth tellers? This behavior is required for democratic governments to work in the best interests of voters. We suspect the reason for its absence is that hardly anyone has received the equivalent of TLT and in particular vote training, which is amazing simple. Vote training consists of following the two simple rules described in the lower right of Figure 4.

**H5.** *Training on LTQ and DTQ persists but falls over time.* This was partially supported by second questionnaire results.

**H6.** *The fall in LTQ and DTQ over time may be eliminated with sufficient refresh training.* This was weakly supported by second questionnaire results. However, we expect the problems causing weak results can be eliminated as discussed earlier.

In all cases, replication and further research is required to more fully confirm, reject, or modify these hypotheses. Study results apply only to the fallacies subjects were trained on, and are thus too narrow to apply to real world behavior. A broader solution will require broader training and much further development.
### Vote question results

<table>
<thead>
<tr>
<th>Treatment Groups</th>
<th>Deceptive statements</th>
<th>Non-deceptive statements</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Trained on neutral topic (control group)</td>
<td><img src="image" alt="Graph A" /></td>
<td><img src="image" alt="Graph B" /></td>
</tr>
<tr>
<td>2 Trained on claims</td>
<td><img src="image" alt="Graph C" /></td>
<td><img src="image" alt="Graph D" /></td>
</tr>
<tr>
<td>3 Trained on claims and vote</td>
<td><img src="image" alt="Graph E" /></td>
<td><img src="image" alt="Graph F" /></td>
</tr>
</tbody>
</table>

#### The Vote Question

If the election were held today and this was all the information you had, how much impact would what the politician claimed have on your decision to vote for or against the politician?

#### Answers

1. Very large increase in support.
2. Large increase in support.
3. Medium increase in support.
4. Small increase in support.
5. It would make no difference.
6. Small increase in opposition.
7. Medium increase in opposition.
8. Large increase in opposition.
9. Very large increase in opposition.

### Figure 7

Distributions of the vote question answers for the first questionnaire. The correct answer is 9 for deceptive and 1 for non-deceptive statements. A normal distribution curve was added to chart A for discussion.

The vote question answers in Figure 7 show interesting behavior:

**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. Note how close the data comes to a normal distribution centered on the midpoint. This indicates a person’s level of political truth literacy is largely due to random factors (environmental and genetic chance) rather than the formal education seen in charts C and E.
Chart B shows untrained people lean toward supporting truth telling politician, but seem shy about supporting them strongly with answer 1. We offer no theory why this is so.

**Group 2. Trained on claims** – As discussed earlier in H4, the second row contains what to us is astounding 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%) chose 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 too is puzzling behavior we cannot explain.

**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 from low to high. The analysis paper concluded that a range of 80% to 90% for LTQ and AAQ is required for satisfactory root cause resolution, though much further research is necessary. DTQ = LLQ x AAQ, so this equals a range of 64% to 81% for DTQ, which appears feasible.

To accelerate solution implementation, we estimate only 10% to 20% of an electorate needs training since most elections are close. The biggest training impact would be on uncommitted young and swing voters. Voters already strongly committed to a false ideology will tend to resist change due to the deceptive power of motivated reasoning, as discussed later. Training is not urgently needed for voters already supporting truth-telling politicians. This suggests that initially, training should target those who would benefit the most. In the long-term, all citizens should be trained.

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.

**Correct answers**

Political deception has become so destructive in today’s world that nothing less than answer 9 for deceptive statements and answer 1 for non-deceptive statements is required for democracy to thrive. Electing truth-telling leaders (in terms of the objective truth about what’s best to optimize the common good) must be the top priority of voters.

For deceptive or true statements, even small deviation from the correct answer matters, since that indicates a person has been partially deceived and that adds up. Koch and Arendt surveyed the research on cumulative media effects and found that “small effects may become decisive and meaningful when they accumulate over time” due to repeated exposure.

Deception involves three components: (1) *Response shaping,* the initial formation of a new slightly different response to a stimulus, such as a slight amount of
doubt climate change is real, a small dislike of immigrants, some doubt a politician not a liar, or skepticism to wearing face masks during the Covid19 pandemic. (2) Response reinforcement, where each repetition of a falsehood (often using different false evidence) strengthens the false belief response. (3) Response change, how a person responds to a stimulus based on the amount of false belief acquired.

The cumulative exposure effect occurs because of component 2. Each additional exposure (even if the deceptive effect is small, like answer 4 to deceptive statements) reinforces the conditioned response. “The strength of people’s attitudes depends entirely on the number of incoming messages about the attitude issue they have processed” (p76).

Consider deceptive statements. What begins as a small error in reasoning, such as a small deviation from the correct answer, with many repetitions grows into a large error, such as reasoning that “Well, Politician A didn’t tell that big a lie. I’ll vote for him anyway.” Or they might reason “Just because Politician B denies climate change is real is not that strong a reason to oppose him. Other things matter more.”

For non-deceptive statements, small deviation from the correct answer also matters, since that indicates a person does not understand how to best support truth-telling politicians. These errors accumulate and can lead to large errors, such as: “I like the fact Politician C told the truth on such an important subject. But telling the truth is not that big a deal. Other factors like experience are just as important.”

**Nullifying the deceptive power of motivated reasoning with high-speed pattern recognition**

The analysis found the high leverage point for resolving the main root cause of democratic backsliding to be raise political truth literacy from low to high. The study offers empirical evidence political truth literacy is presently low and can be fairly easily raised to high, though this was a laboratory study rather than an actual democratic system.

The cognitive mechanism employed in the training is high-speed pattern recognition by spotting patterns of non-fallacies (truth) or fallacies (deception). Subjects were trained on how to spot one pattern of the truth: correct application of the Strong Evidence Rule. This inductive rule of logic appears to be the most common rule in true political statements. Subjects were also trained to spot flawed application of the Strong Evidence Rule plus six fallacies common in political falsehoods: cherry picking, ad hominem attack, appeal to emotion, strawman, false dilemma, and false fact lie.

A high-seed pattern recognition approach can nullify the deceptive power of motivated reasoning, a well-established theory explaining how biased decision-making works. The theory explains why once a person is fooled into believing deceptive goals and facts, the person becomes highly partisan and their false beliefs are unshakable. Instead of thinking logically, they behave as The Rationalizing Voter, the title of Lodge and Taber’s magnum opus, which summarizes decades of empirical research. “In short, citizens are often partisan in their political information processing, motivated more by desire to maintain prior beliefs and feelings than by their desire to make ‘accurate’ or otherwise optimal decisions”.

When a prior belief is false (such as non-
whites are inferior or climate change denial), deception has occurred and partisan reasoning will be erroneous. The research centers on the John Q. Public simulation model28, a calibrated node-link associative model, which “has been remarkably successful… in replicating the processing and behavior of real citizens”29.

The purpose of deception is to create a false belief that benefits the deceiver at the expense of the deceived. Motivated reasoning theory divides false beliefs into two broad types: False goals and false facts (or claims) used to rationalize false goals. In a democracy, false goals are those that do not benefit the common good (the majority of the electorate), and benefit special interests (a minority) instead.

False beliefs are created and strengthened by fallacious arguments, which work when someone fails to spot a fallacy. But once a person learns the spot-the-pattern of truth or deception technique of TLT, they are inoculated. Fallacies they have been trained on or logic they cannot identify can no longer be used to fool them into a false belief, so they never believe the false belief in the first place or may question a false belief already held. This depends on the level of truth literacy a person has attained.

The premise of motivated reasoning theory is that all reasoning is motivated to achieve either accuracy goals (slow thinking) or partisan goals (directional, fast thinking). With enough training and experience in how to “read” the truth by pattern recognition (claim training), and how to use that knowledge to act correctly (vote training), sufficiently correct accuracy reasoning can approach the speed of partisan reasoning and replace it, thereby becoming the reasoning default when important new political arguments or facts are encountered, or old ones need review.

“…the ways by which we are deceived are consistent and not so hard to recognize”.30 A thorough program of pattern recognition training can create a reliable high-speed pattern recognition heuristic. With proper training, accuracy reasoning is now automatically used instead of partisan reasoning when confronted with new potentially false inputs, because accuracy reasoning is now fast instead of slow, and usually correct instead of so easily deceived. A person’s important political beliefs will now tend to be true instead of false, depending on their level of truth literacy.

Heuristics work by substituting fast and frugal reasoning for slower logical reasoning. “A heuristic is a strategy that ignores part of the information, with the goal of making decisions more quickly, frugally, and/or accurately than more complex methods”.31 Pattern recognition allows irrelevant information to be ignored and not processed, which can speed reasoning so much it becomes instantaneous.

For example, the statement in Figure 4 contains 85 words about a Trade Agreement Treaty. As soon as a truth literate person spots the false dilemma pattern in the 6 words saying “There are only two choices here” and confirms there really are more than two choices (there almost always are in politics), they know the statement is fallacious and nothing else matters. Step 3 of the Personal Truth Test (Figure 4, right panel) applies.

TLT employs the preemptive aspect of inoculation theory. Innovating by training on logic pattern recognition instead of misinformation correction (such as fact-checks and news pointing out the truth), as we have done, Cook et al.32 found that inoculating subjects by training on spotting false balance and fake expert strategies
“neutralized” the negative influence of misinformation on perceived scientific consensus on climate change.

Our approach necessarily goes one step further by introducing vote training and greatly improves training effectiveness by adding a catalog of common fallacies and The Personal Truth Test. Drilling subjects on the catalog of common fallacies (and correct and flawed application of the Strong Evidence Rule) teaches high-speed pattern recognition in the same manner that students are drilled on letters of the alphabet, words, numbers, multiplication tables, object names from pictures of a cat, dog, etc. With sufficient practice citizens can now “read” the truth. Common fallacies can be spotted in seconds and false meme infection prevented most of the time.

Peering into the future, a society that has achieved universal political truth literacy might look something like this:

Marketplace Deception Protection Skills…. A person who is skilled in deception protection will have well-learned mental procedures designed to detect, neutralize, resist, correct for, and penalize deception attempts…. More broadly, consumers adept at deception self-protection will learn to warn and protect friends, kin, and loved ones…. Most broadly, consumers must learn to adopt a deception protection goal as their default….33

Additional solution elements

The larger research project designed a collection of solution elements for pushing on the high leverage point. The most promising one to develop first was TLT, since it re-quires the least amount of work to develop, test, and implement for the highest impact, and is the core solution element.

Six additional solution elements were designed. The first provides the governance foundation and incentive for the rest. The second provides the legal foundation for in-formation availability. The third and fourth are ratings, which provide the truth about politician behavior. The fifth and sixth are indexes, which provide the truth about the state of the world (or the local system) that politicians are managing for the public good. Ratings and indexes supplement TLT by providing citizens with additional information for decision-making. The additional solution elements are:

1. Freedom from Falsehood – This gives people the legal right to freedom from falsehood from sources they must be able to trust, such as politicians. Just as freedom of speech does not apply to shouting “Fire!” in a crowded theater and causing a panic, freedom of speech does not apply to falsehoods that endanger the wellbeing of democracy. For example, 76% of UK voters support making deliberate lying by politicians illegal.34 Harrar35 explains why:

A lying virus has spread across the body politic. It has always been there, but a new, more toxic variant is threatening democratic life. … If a teacher or a doctor lies to us, there are consequences. Any product brought to the market must comply with a raft of legislation ensuring the seller tells the truth about it—the more potent the product, the greater the duty. The same should apply to politicians.
2. No Public Servant Secrets – A public servant is any entity that exists to serve the people, such as politicians and publicly created artificial life forms like corporations and governmental agencies. Ensuring the accountability and integrity of public servants must be possible at all times. If public servants can keep certain information secret, including conversations, documents, emails, phone calls, texts, and financial transactions, then Freedom from Falsehood cannot be implemented because in too many cases there would be no way to discern the truth. National security secrecy is not exempt, since that form of secrecy encourages war, which reduces quality of life.

3. Politician Truth Ratings – This measures the average truth of important statements made by important politicians. It would be available online in formats like Figure 8. Speeches, ads, articles, and so on are sampled on a running basis to calculate a statistically valid rating. This may seem like an expensive burden, but most claims are repeated. Only the first occurrence requires new work. Until natural language processing and AI can do most of the work, the ratings would be limited to top politicians.

When citizens base voting decisions on truthfulness, they are essentially developing their own Politician Truth Rating for each politician on the ballot. This cannot be done efficiently and accurately by each citizen for every politician of interest due to the time and expertise required. Politician Truth Ratings fill that gap. Ratings of other types have long filled similar gaps, such as credit ratings, car ratings, bond ratings, and school quality ratings.

![Politician A – Truth Ratings](image)

Figure 8. Example of online Politician Truth Ratings. The promise rating addresses the common problem of politicians making false promises to get elected and then failing to deliver what was promised, and supplements the truth rating.

4. Politician Performance Ratings – Like the cumulative result of annual employee performance reviews, this measures how well a politician has performed in
his or her career. For legislative work, the ratings measure a politician’s contribution to problem selection, bill design, voting for or against a bill, and then how well the bill solved the problem. For executive work like governor or prime minister, the ratings measure how well they achieved standing objectives and handled new problems.

5. Quality of Life Index – Maximizing short-term economic growth has long been the implicit goal of “practically all nations” and “the main objective of politics.” The right goal is the long-term optimization of the common good, as measured by the median (not the average) quality of life for a population. Numerous indexes have appeared to do this, such as the Genuine Progress Indicator, the Economist’s Quality of Life Index, the Legatum Prosperity Index, the United Nations Human Development Index, and Bhutan’s Gross National Happiness Index.

6. Sustainable Management Index – The top problem facing humanity today is global sustainability. The index would be a weighted average of all three pillars of sustainability: environmental, economic, and social. More destructive problems, like war, climate change, and high inequality of wealth/income, would receive higher weights. Users could drill down to see the index for particular problems of interest.

Democratic backsliding is arguably the world’s top political problem, since special interests (and authoritarians in particular) care little about common good problems. Ratings and indexes allow citizens to hold politicians accountable for their individual contribution to optimizing quality of life and solving sustainability problems, by serving as supplementary information when making voting decisions.

Figure 9. High-level feedback loop structure of how the seven solution elements (in green) work together to raise DTQ. TLT is the core solution element, since all other solution elements are designed to optimize the effect of TLT. The Self-Improvement of the Common Good loop is discussed in the analysis paper. Solid arrows are a direct relationship. Dashed arrows are an inverse relationship. R and B signify reinforcing and balancing loops.
The solution elements would work together to raise DTQ in a rapid and efficient manner as shown in Figure 9, ideally before more nations tip into irreversible states of authoritarianism, as have China and Russia. Once the Race to the Bottom reaches its authoritarian extreme in a political system, a tipping point is reached and reversal of feedback loop dominance becomes difficult or impossible, due to state control of education and information flow.

Resolving root cause forces requires engineering a feedback loop structure capable of adequately pushing on the high leverage points in a self-improving solution manner. If any of the loops in Figure 9 are weak or any solution elements are missing, the overall solution will fail because the Growth of the Common Good loop would be too weak to raise DTQ to the high level needed to fully resolve the root cause. Further solution elements will be needed, but these seven (or something similar) appear to be the approximate minimum set for realistically solving the backsliding problem. Removing any single element destroys the efficiency of the feedback loops.

Conclusions

The analysis and TLT study provide the first version of a comprehensive theory explaining democratic backsliding. Using a social force diagram (Figure 2) paired with a system dynamics model (Figure 3), the Dueling Loops theory explains:

1. Why democratic systems are backsliding. This is due to the Dueling Loops structure and an unresolved main root cause (force R in Figure 2). Existence of the root cause was verified on a preliminary basis by the TLT study.
2. Why popular solutions tend to fail. Superficial solutions (force S) have pushed on intuitively attractive low leverage points.
3. Why certain innovative solutions can be expected to succeed. The TLT study provides preliminary evidence that fundamental solutions (force F) can push on the correct high leverage point in order to resolve the main root cause.

We argue that for a theory explaining a difficult social problem to be sufficiently complete, it must explain all three forces (R, S, and F) and include sufficient proof of all key assumptions.

The key analysis assumption is the main root cause and its high leverage point: raise political truth literacy from low to high. This assumption was why the TLT study was performed. But this raises a question. How can a notoriously complex problem like democratic backsliding have such a simple solution strategy? Doesn’t this indicate the analysis is somehow overly simplistic and flawed? We think not for two main reasons:

1. 100% of authoritarians (as well as all far-right parties and politicians) depend on copious amounts of propaganda/deception to fool voters into supporting them. If political truth literacy was high instead of low, this strategy would no longer work.
Hannah Arendt, in her seminal work on The Origins of Totalitarianism, reached essentially the same conclusion. The work addressed the fundamental question, concerning Russian Stalinism and German Nazism, of “Why did it happen?” The main precondition was found to be susceptibility to propaganda:

The masses have to be won by propaganda. ... The ideal subject of totalitarian rule is not the convinced Nazi or the convinced Communist, but people for whom the distinction between fact and fiction (i.e., the reality of experience) and the distinction between true and false (i.e., the standards of thought) no longer exist.

Inability to distinguish between fact and fiction, true and false, in political decisions is the same as low political truth literacy. Once the main root cause of low political truth literacy is resolved by pushing on the high leverage point, the “ideal subject of totalitarian rule” would no longer exist in sufficient quantity for the path to totalitarianism to work.

Puddington, in a study of the methods of modern authoritarian states, found that Russia, more than any other country including China, has refined “the ideas and institutions at the foundation of 21st century authoritarianism.” Under Vladimir Putin, Russia’s most important innovation has been “the development of modern propaganda and disinformation” mechanisms. China and Russia, the world’s two major authoritarian powers, employ propaganda as the central strategy for maintaining popular support and justifying any actions leadership may choose to take.

2. The high leverage point has never been pushed on before with focused, large-scale, long-term solution elements:

(a) No education system has ever deeply educated students in political truth literacy, right along with reading, writing, and math literacy.
(b) No large news organization, public or private, has ever made continuing education of the public in political truth literacy part of its mission, via news coverage that include components of TLT. One example would be a news story on how Hitler, Vladimir Putin, and Donald Trump each employed the classic authoritarian deception pattern of painting a false common enemy to push the fear hot button and justify violence against that enemy, when in reality no such enemy existed. Or there might be a daily series on “The smallest small lie, the biggest big lie, and the most novel fallacy of the day.” The point of this info-education, instead of infotainment, is spot the pattern and don’t be fooled.
(c) No major research project has ever centered on developing empirically based approaches to TLT and other solution elements for pushing on the high leverage point.
(d) No government has ever adopted the policy of universal political truth literacy. Yet we have long witnessed the extraordinary gains from universal reading, writing, and math literacy. If the Dueling Loops theory is sound, then we can expect the same extraordinary gains from universal political truth literacy.
Some education systems have offered training in skills related to truth literacy, like critical thinking. Definitions of critical thinking vary widely. Robert Ennis, one of the founding fathers of the critical thinking movement in North America, offers this definition: “Critical thinking is the intellectually disciplined process of …evaluating information …as a guide to belief and action.” This describes the ability to correctly reason in general. Critical thinking is not the same as the specific skill of political truth literacy. Nor is media literacy.

We look forward to the next phase in the long evolution of governance, when due to universal political truth literacy, a permanent greatly strengthened Race to the Top among Politicians has begun and is on its way to spreading to all nations. Politicians now cooperatively compete on the truth about what’s best for their constituency, instead of destructively competing to see who can deceive the most voters into voting against their own best interests. The universal goal of governments and politicians would be to optimize quality of life for the common good of all, for those living today and all who come after. This, of course, is the ultimate goal of democracy.

Notes

1 Francis Fukuyama, The End of History and the Last Man.
2 Arsel, Adaman, and Saad-Filho, “Authoritarian Developmentalism.”
3 Bermeo, “On Democratic Backsliding.”
4 Waldner and Lust, “Unwelcome Change.”
5 Wiesner et al., “Stability of Democracies.”
7 Luhmann and Lindberg, “A Third Wave of Autocratization Is Here.”
8 See Redlawsk, David, “Voting Correctly.” The paper introduced the behaviour rule that voters seek to vote “correctly” from their own (and not a normative) perspective, and introduced a method for measuring voting correctness. Redlawsk concludes that most of the time, voters vote correctly. However, we point out this assumes that their decision-making information was not deceptive. If that information is deceptive, then voters will think they are voting correctly, when if fact they are not. That’s why the definition of political truth literacy includes the phrase “logically correct voting decisions.”
11 Popper, All Life Is Problem Solving, 13, italics in the original.
12 Svolik, “Polarization versus Democracy.”
13 Levitsky and Ziblatt, How Democracies Die, 5.
14 Foa, “Modernization and Authoritarianism,” 133.
16 Noel and Therueb, Left and Right in Global Politics, 3.
17 Larrabee, Bentham’s Handbook of Political Fallacies, xxi.
19 Sagarin et al., “Dispelling the Illusion of Invulnerability: The Motivations and Mechanisms of
Resistance to Persuasion.”
21 Gliem and Gliem, “Calculating, Interpreting, and Reporting Cronbach’s Alpha Reliability Coefficient for Likert-Type Scales.”
22 Taber, “The Use of Cronbach’s Alpha When Developing and Reporting Research Instruments in Science Education,” 1287.
23 Koch, Arendt, and Maximilian, “Media Effects: Cumulation and Duration.”
25 Kunda, “The Case for Motivated Reasoning.”
26 Lodge and Taber, *The Rationalizing Voter*.
27 Ibid p149.
29 Ibid p204.
32 Cook, Lewandowsky, and Ecker, “Neutralizing Misinformation through Inoculation: Exposing Misleading Argumentation Techniques Reduces Their Influence.”
34 Glaze, Ben, “Lying by Politicians Should Be Made a Specific Criminal Offence, Say Voters.”
35 Harrar, Jaya, “Can the Law Sort out Political Lies?”
36 *2052: A Global Forecast for the Next Forty Years*, 73.
38 Arendt, *The Origins of Totalitarianism: Part Three*.
40 Ibid, p172.
42 Ennis, “Critical Thinking: A Streamlined Companion.”

Acknowledgements

(To be added later.)

Disclosure statement

The authors report there are no competing interests to declare.

Bibliography

Anderson, Bjorn, and Tom Fagerhaug. *Root Cause Analysis: Simplified Tools and*


Levitsky, Steven, and Daniel Ziblatt. How Democracies Die. Crown Publishing Group,
2018.


