Cutting through the complexity of the democratic backsliding problem with root cause analysis

Jack Harich (1) and Montserrat Koloffon Rosas (2)
(1) Thwink.org, Atlanta, USA, jack@thwink.org (corresponding author)
(2) Vrije Universiteit Amsterdam, Institute for Environmental Studies
Manuscript Version: November 17, 2023

Appendix

Contents

The simulation model

How system dynamics models work

For simple problems an essential causal structure (ECS) diagram alone is sufficient. For problems where complexity hides the feedback loop structure of any portion of the diagram (especially the fundamental layer), an additional root cause analysis (RCA) sub-tool is required: feedback loop simulation modeling.

The most direct form of this tool is system dynamics, due to easy-to-use software (like Vensim and Stella) allowing visual portrayal of feedback loop structure and running of scenarios. System dynamics is a modeling language for understanding how feedback loop structure causes a problem's emergent behaviour and how that behaviour can be modified, and is widely used on complex business problems (Sterman, 2000). System dynamics can be supplemented or even replaced by other forms of cause-and-effect modeling, such as agent-based, as long as the essential feedback loop structure is identified and understood.

The fundamental principle of system dynamics states: "The behaviour of a system arises from its structure. That structure consists of the feedback loops, stocks and flows, and nonlinearities created by the interaction of the physical and institutional structure of the system with the decision-making processes of the agents acting within it" (Sterman, 2000, p. 107). It follows that if problem solvers don't understand a system's feedback loop structure, then they don't understand the system. Solution of highly complex problems will be impossible, except through long trial and error, and occasional luck.

The Dueling Loops of the Political Powerplace model (Figures 2 and 3) follows the long tradition of using small system dynamics models to find and communicate powerful insights (often counterintuitive) to the public and policy makers, (Ghaffarzadegan et al., 2011) such as Jay Forrester's iconic World2 model with its 56 variables (Forrester, 1971, pp. 20–21). Because the Dueling Loops model (41 variables including those not shown) is relatively easily understood and exhibits clear behaviour, "important insights regarding the source of policy failures can be uncovered" (Ghaffarzadegan et al., 2011, p. 36). The ultimate source is the unresolved main root cause.

Small insight models entail estimated parameters for archetypical/exploratory use or as many measured parameters as feasible for actual cases. The Dueling Loops of model was based almost entirely on estimates. After the Truth Literacy Training study was complete, we were able to calibrate the LTQ and AAQ nodes. The values used are described later along with the simulation runs. Using modeler judgement based on system observation plus calibrating the high leverage points with study results, the Dueling Loops model was tuned to give realistic behaviour over the full range of the high leverage points, *false meme size*, and *influence per degenerate* or *rationalist*.

Note the simplicity of the model. Only a few dozen nodes are required to mimic the relevant behaviour of the system from a root cause analysis point of view, despite the fact that the democratic backsliding problem consists of the behaviour of billions of people, 195 countries, and thousands of head-of-state politicians over the history of the problem. Such

simplicity is possible only with an appropriate problem-solving process that allows cutting through the complexity of the problem and reducing it to its essential causal structure.

Model simplicity provides substantial benefits. With only a modest amount of attentive study, the simulation model described here can be understood and verified by a wide range of scholars, not just modeling specialists. A simple model is much easier for other modelers to deeply understand and evolve as the analysis undergoes further iterations or as the model is applied to other problems. The benefits of a small simple model are so large we have deliberately kept the model as small and as clearly organized as possible.

Small insight models use estimated parameters for archetypical/exploratory cases or as many measured parameters as feasible for actual cases. The Dueling Loops of model was based almost entirely on estimates. After the Truth Literacy Training study was complete, we were able to calibrate the LTQ and AAQ nodes. The values used are described later along with the simulation runs. Using modeler judgement based on system observation plus calibrating the high leverage points with study results, the Dueling Loops model was tuned to give realistic behaviour over the full range of the *high leverage points, false meme size,* and *influence per degenerate* or *rationalist*.

A model is *calibrated* by measuring factors in the real world, using those values in the model, and running the model to compare its dynamic behaviour to that in the real world. Model structure is then improved as needed to narrow the gap between model and real-world behaviour to an acceptable level.

System dynamics is a simulation modeling language that models the structure of a system in terms of its feedback loops and stocks, and how that structure causes behaviour change over time. The goal of system dynamics is "to enhance learning in complex systems… to understand the source of policy resistance, and design more effective policies" (Sterman, 2000, p. 4).

System dynamics modeling approximates a problem's behaviour, by comparing graphs of model behaviour to graphs of collected data in the real world and refining the model until graph agreement is good enough. However, the main purpose is to generate useful insights by understanding the problem's feedback loop structure, so that the analyst knows WHY various problem behaviour occurs. If a model's general behaviour resembles problem behaviour, and model structure makes complete sense and corresponds to the real world, then it is a useful model, whether it is a small simple insight model or a larger more complex calibrated model.

Calibration adds considerable model complexity. For example, Forrester's (1971) World2 system dynamics model used a mixture of estimated and measured parameters, and was an insight model. Its successor, World3 (Meadows et al., 1972, 1974), was fully calibrated. This increased total number of variables from 56 to about 320.

How feedback loops work

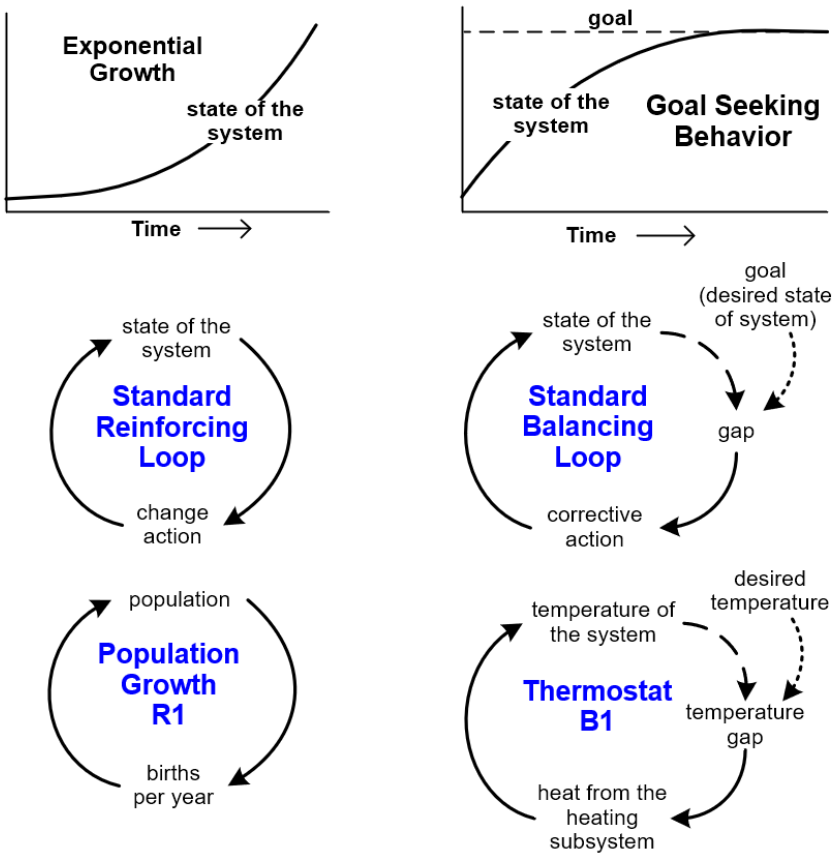


Figure 1. Examples of reinforcing and balancing loops and their graphs.

Feedback loops (Figure 1) control the behaviour of a system over time, as shown in graphs. **Reinforcing loops** cause either runaway exponential growth (as shown) or exponential decay (not shown). In the **Standard Reinforcing Loop**, the *state of the system* starts at some non-zero amount. This state causes a *change action*, which in turn increases the *state of the system*, which in turn increases the *change action* even more. The loop grows exponentially due to its self-reinforcing nature. The graph illustrates the resulting exponential growth.

A real-world example of a reinforcing loop is **Population Growth**. As *population* goes up, so does *births per year*. As that goes up, so does future *population*. This increases *births per year* still further. The loop goes round and round, growing exponentially until the loop hits its limits to growth, which are not shown.

The other type of loop is a balancing loop. **Balancing loops**, also known as goal seeking loops, have a goal. The loop "balances" its behaviour so as to reach the goal. In social systems they do this by allowing intelligent agents to measure the gap between what is and what's wanted (which is the goal) and make decisions so as to close the gap.

In the **Standard Balancing Loop** the *gap* equals the *goal* minus the current *state of the system*. Suppose the loop starts with a low *state of the system* as shown in the graph. The *goal* is much higher, so the system starts with a high *gap*. The greater the *gap*, the larger the *corrective action*. As *corrective action* increases, so does the *state of the system*.

Here's where balancing loops differ from reinforcing loops. As the *state of the system* goes up, the *gap* goes down. This reduces the *corrective action*. The loop continues until the *gap* is zero, at which point the loop has reached its *goal* and further change ends. The graph illustrates this behaviour.

A common example of a balancing loop is a **Thermostat**. Suppose you set the *desired temperature* to 70 degrees Fahrenheit but the current *temperature of the system* is 65. That causes a *temperature gap* of 5 degrees. The greater the gap the greater the *heat from the heating subsystem* that flows into the system. This increases the *temperature of the system*. As this goes up the *temperature gap* goes down. The loop goes round and round until the gap is zero, at which point the thermostat system has reached the goal. (The simple example assumes the heating subsystem is capable of variable output instead of on/off behaviour.)

How the simulation model works

The main paper explained how the causal loop diagram worked. The fundamental layer of that diagram (as well as intermediate cause 4) was a high-level summary of the system dynamics simulation model (Figure 2). The model contains many more nodes (variables) in order to simulate the model.

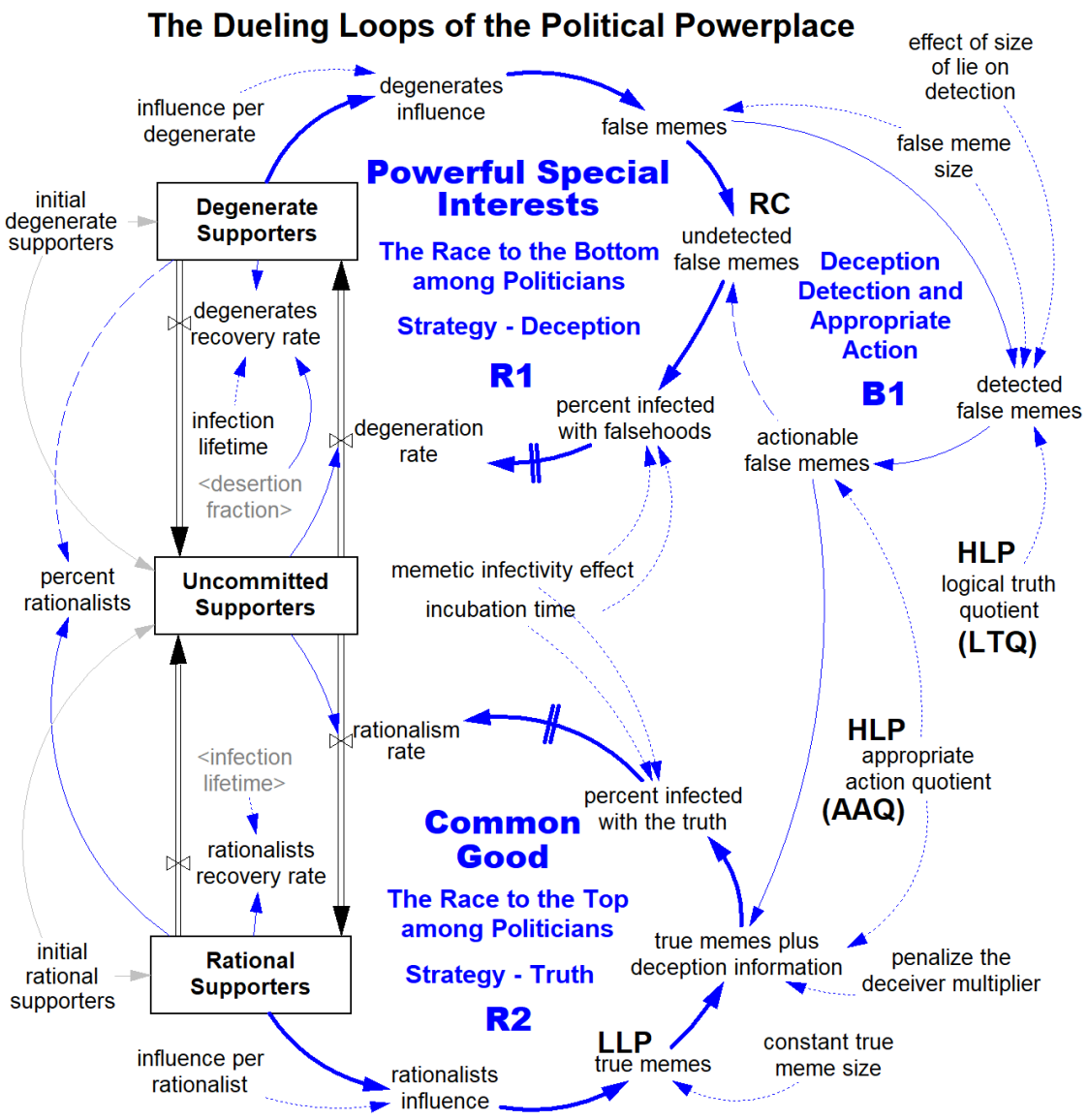


Figure 2. System dynamics model of the democratic backsliding problem. Unit conversion nodes like *one dollar* and *one year* are not shown, as these have no effect on model behaviour. Legend: As in the causal loop diagram in the paper, an arrow from node X to node Y means X causes change in Y. Solid arrows are a direct relationship, meaning as X increases so does Y, or as X decreases so does Y. Dashed arrows are an inverse relationship, meaning as X increases Y decreases and vice versa. Dotted or light grey arrows are constants, indicating X remains constant. R and B signify reinforcing and balancing loops.

The three boxes are what system dynamics calls "stocks." Stocks represent the most important factors whose behaviour you are trying to understand and usually correspond to physical objects, such as people, money, and pollution. Stocks form the backbone of a system dynamics model. Objects flow between stocks according to "rates." The model uses four rates to move supporters from one stock to another.

The model uses Dawkins' concept of memes and memetic replication (Dawkins, 1976, pp. 189–201). A meme is copied information capable of affecting behaviour. All memes are

learned from others, either directly from other people or indirectly through a transmission medium such as books, television, or social media. Replication is also called memetic infection. A person is infected by a meme when it enters and is accepted by their mind. In the model a meme is a statement that is true or false.

The model uses the concept of memetic infection to determine the *degeneration rate* and the *rationalism rate*. Looking at the Race to the Bottom, *undetected false memes* is used to calculate *percent infected with falsehoods*. After a delay of 1 year,[1] which is *incubation time*, the infection matures enough to cause the *degeneration rate*. This causes *Uncommitted Supporters* to move to the *Degenerate Supporters* stock. The Race to the Top works in the same manner.

People don't stay infected forever. Some eventually recover. The model handles this with the two *recovery rate* nodes. Average length of infection is 30 years, which is *infection lifetime*. This causes 3.3% (1/30 = .033) of those in the *Degenerate Supporters* and *Rational Supporters* stocks to recover each year and move back to the *Uncommitted Supporters* stock. The 30 years is estimated, using the fact that people rarely switch parties (*Voters Rarely Switch Parties, but Recent Shifts Further Educational, Racial Divergence*, 2020).

All this mimics what we see in the real world. People are exposed to a mixture of *true* and *false memes* via TV, social media, articles, books, conversations, etc. Depending on their LTQ and AAQ, some of the *false memes* become *actionable false memes (AFM)*. This causes three things:

1. AFM is subtracted from *false memes* to calculate *undetected false memes*.
2. AFM is added to *true memes* to calculate *true memes plus actionable false memes*.
3. AFM causes some degenerates to see the truth and desert. AFM is used to calculate *desertion fraction* (Figure 3). This and *infection lifetime* are then used to calculate the *degenerates recovery rate*. If the *desertion fraction* is greater than zero, this increases the recovery rate.

Two constants define the difference between the two main loops, which otherwise are equal in their attractive power. In the Race to the Top, *constant true meme size* is always one. It can never change, because the attractiveness of the truth cannot be inflated. But in the Race to the Bottom, *false meme size* ranges from one and up, because the attractive power of a meme can be inflated with deception. This constant is changed to different values in the simulation runs. Each change represents an optimum deception strategy: "How much lying can I get away with to maximize the number of my supporters?"

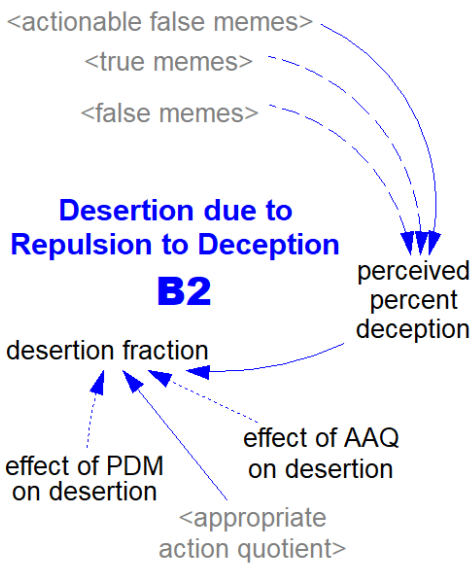


Figure 3. Variables used to calculate the *desertion fraction*. This forms part of a fourth important feedback loop that's not on the main model.

The ability to inflate the attractiveness of a meme gives the Race to the Bottom an inherent advantage, represented by *undetected false memes*. The Race to the Top has no corresponding node, like *undetected true meme*, since there is no deception to detect in the truth. This advantage is the main root cause of why the Race to the Bottom is dominant most of the time, since LTQ and AAQ are presently low.

For simplicity, we usually say the main root cause is low political truth literacy. Political truth literacy is DTQ. DTQ equals LTQ times AAQ.

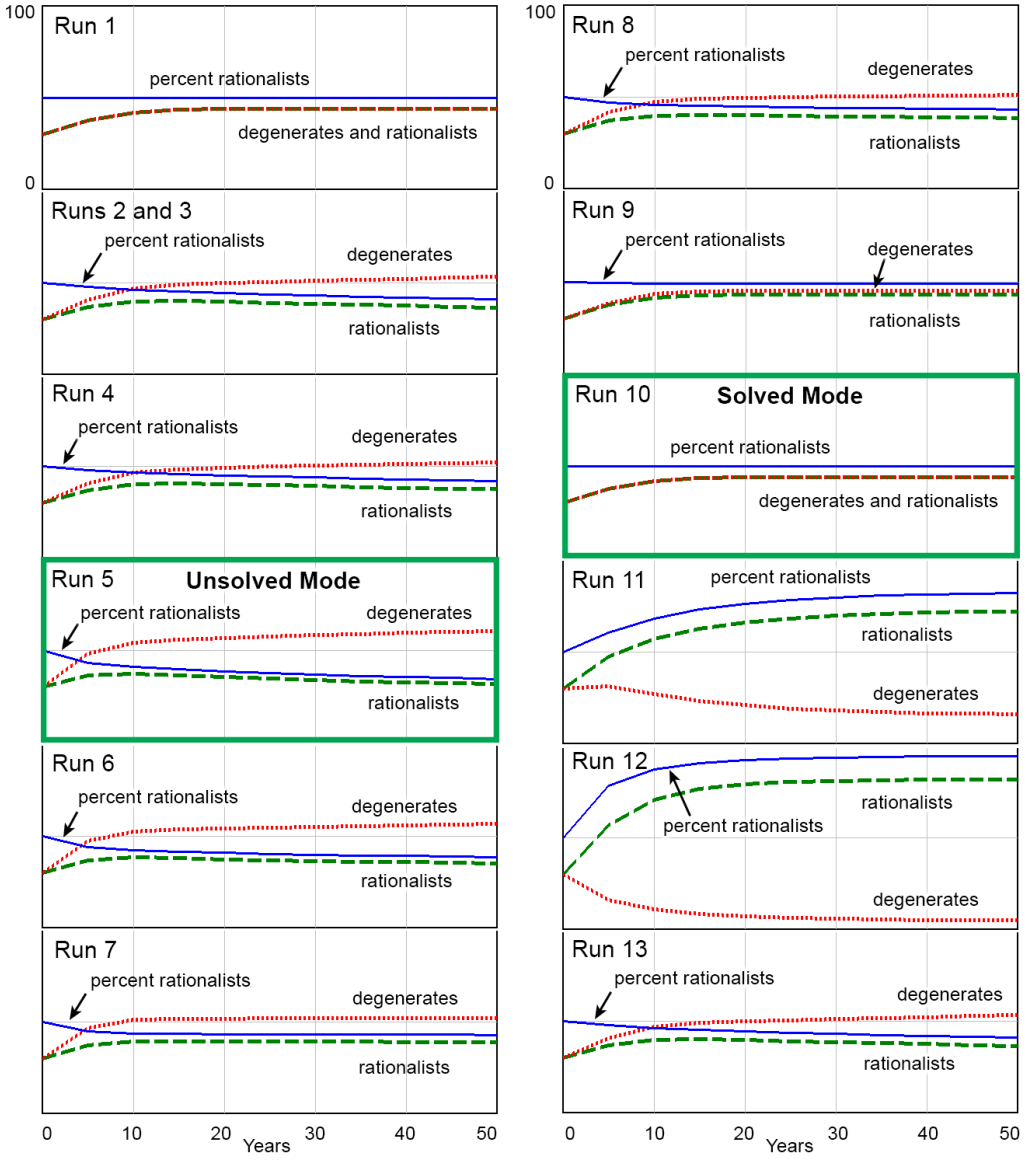
For further detail, the model considers influence. In the Race to the Bottom loop, *influence per degenerate* times the number of *Degenerate Supporters* equals *degenerates influence*. For simplicity, one unit of influence equal one *false meme*. The Race to the Top has corresponding nodes.

The simulation model is a reasonable approximation of how the root cause, low leverage point, and high leverage point in the ECS diagram work. Without the simulation model, it would have been *impossible* to correctly explain the superficial and fundamental layers of the ECS diagram. We would have never found what appears to be the main root cause and its two high leverage points.

As simple as the Dueling Loops model looks, construction of the first version took about three years. When we began, we had no idea what would be found on the fundamental layer of the problem. Fortunately, we had the guiding hand of root cause analysis and eventually identified the basic feedback loop structure that appears to exist in all large political systems.

Simulation runs

This section shows how the model behaves using a series of simulation runs.



Model Settings	Simulation Runs Run 5 is the problem to solve. Run 10 is the solved mode.												
	1	2	3	4	5	6	7	8	9	10	11	12	13
False meme size	1	1.5	1.5	1.5	4.9	4.8	3.4	2	1.3	1	4.9	4.9	1
LTQ – Logical truth quotient	0	0	NA	20%	20%	30%	20%	30%	40%	50%	50%	70%	50%
AAQ – Appropriate action quotient	0	NA	0	20%	20%	20%	30%	30%	40%	50%	50%	70%	50%
DTQ – Democratic truth quotient	0	0	0	4%	4%	6%	6%	9%	16%	25%	25%	49%	25%
Influence per degenerate	200	200	200	200	200	200	200	200	200	200	200	200	300
Results													
Percent rationalists at end of run	50%	41%	41%	42%	34%	38%	42%	43%	49%	50%	82%	94%	41%

Figure 4. Simulation run graphs, settings, and results. Before a run, the five settings are set to the values shown. The model is then simulated. Results are then measured. The model is so simple it starts in one equilibrium and moves smoothly to a second equilibrium.

Each model simulation run is a logical experiment. The result tells us how certain things in the real world can be expected to behave, given particular starting conditions. This form of experimentation is orders of magnitude faster and cheaper than real-world experimentation. This explains why system dynamics modeling is so useful, not just on business problems (where the tool was born), but in the social sciences where real-world experimentation is slow, expensive, or impossible.

Experiment inputs are the constants changed. These are *false meme size,* LTQ and AAQ. While listed as a model setting, DTQ is not on the model but is calculated in the Figure 4 table of simulation runs for greater understanding. DTQ equals LTQ times AAQ. The output of interest is *percent rationalists*, calculated by:

percent rationalists = Rational Supporters /
(Rational Supporters + Degenerate Supporters)

Percent rationalists measures dominance of the Race to the Top loop. The higher the percent, the lower democratic backsliding is because fewer people have degenerated.

Figure 4 shows 13 simulation runs. All begin with 40 *Degenerate Supporters*, 40 *Rational Supporters*, and 20 *Uncommitted Supporters*. This gives a total of 100 supporters. In all runs *influence per degenerate* or *rationalist* are equal and never changed. When a simulation run begins, neither side has an advantage except that provided by the model settings.

Run 1. The first run shows how when neither side (rationalists and degenerates) has an advantage, *percent rationalists* stays unchanged at 50%. The number of rationalists and degenerates rise evenly as some neutralists move to the other two stocks. Neither side has an advantage since *false meme size* = 1 and *logical truth quotient* (LTQ) or *appropriate action quotient* (AAQ) equals zero, causing DTQ to be zero.

Run 2. *False meme size* is raised from 1 to 1.5. While telling small lies offers only a small advantage, over time it accumulates into a large one. After 50 years, *percent rationalists* falls to 41%. This is enough for degenerates to win most elections.

In this run LTQ equals zero, so AAQ doesn't matter. If no *false memes* are detected, there can never be any *actionable false memes*, since *false memes* minus *detected false memes* equals *actionable false memes*.

Run 3. This run keeps *false meme size* at 1.5, and instead has AAQ equal to zero. Because of this, LTQ doesn't matter, since AAQ times *detected false memes* equals *actionable false memes*.

As in run 2, DTQ is effectively zero and graph behaviour is identical to run 2. Runs 2 and 3 demonstrate that *both* high leverage points must be pushed on for political truth literacy solutions to work. Let's do that in the next run.

Run 4. Like runs 2 and 3, this run keeps false meme size at 1.5. However, both LTQ and AAQ are set to 20%, causing DTQ to equal 4%. While DTQ is quite low, this is enough to make a small difference. At the end of the run, *percent rationalists* has risen from 41% to 42%.

Run 5. Unsolved Mode. Social agents are adaptive. Degenerate politicians are clever enough to adjust the size of lies to the optimum size: not too big and not too small. The *effect of size of lie on detection* is a lookup table (Figure 5) whose curve reflects how as the size of a lie grows, it's more easily detected and diminishing returns begin. As size rises from 1 to 50, the percent detected rises from zero to 100%. The second curve, *memetic infectivity effect*, is used to calculate *percent infected with falsehoods* and *the truth*.

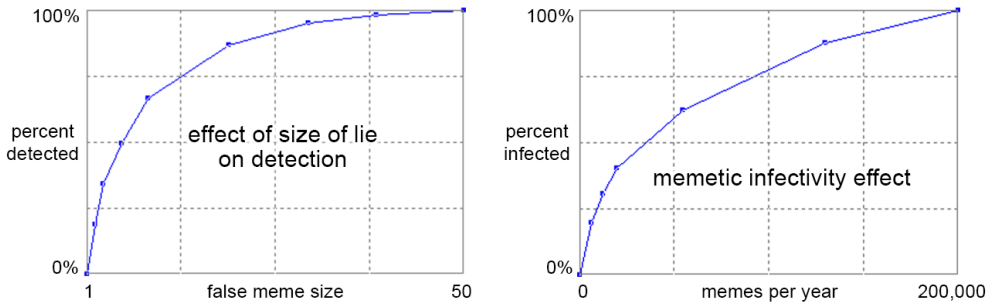


Figure 5. Lookup table values for two curves. Input is the x axis; output is the y axis. Both are non-threshold model curves (non-S curves), very similar to those observed in respiratory disease infection (Sze To & Chao, 2010). The disease curves use probability of infection instead of percent detected or percent infected.

Visual experimentation with the running model shows the optimum *false meme size* is 4.9. Compared to the prior run this gives the degenerates a much larger advantage of 34% *percent rationalists*, which is 66% degenerates.

Based on the Truth Literacy Training study and system observation, we hypothesize that in most political systems both high leverage points (LTQ and AAQ) are low, at about 20%. Run 5 thus reflects approximate real-world behaviour and is what system dynamics calls the reference mode. It is the problem to solve, the run we keep referring back to as we iterate the model and design the other runs.

In run 5 the system is in the unsolved mode. Powerful feedback loop forces are locking the system into a usually dominant Race to the Bottom among Politicians. What's driving these forces is the unresolved main root cause. This is force R in the ECS diagram.

The values of 20% for LTQ and AAQ are calibrated from group 1 (control group) in the study. Group 1 LTQ was 8% in the first study, 22% before the follow up refresh training, and 20% after the refresh training. These average roughly 20%. The AAQ scores were 25%, 59%, and 20%. The 59% is an aberration due to the large confidence intervals and the (accidentally) easier follow up statements. Thus, these also average about 20%.

Using the World Values Survey, Noel and Therien (2008, p. 34) report that 24.7% of respondents place themselves on the left, 30.3% on the center, and 45.2% on the right. This gives a *percent rationalists* of 24.7% / (24.7% + 45.2%) = 35%, which compares favorably to the 34% rationalists in run 5. This is another form of model calibration.

Because political truth literacy is low, the Race to the Bottom is the dominant loop most of the time, where "Special interests now take precedence over the common good. …we now live in a diminished democracy …with ordinary citizens squeezed out of the public sphere

by partisan ideologues and professional propogandists" (Dillard & Shen, 2013, p. 16). The Race to the Bottom is not dominant all the time, because loop dominance changes back and forth due to a variety of reasons beyond the scope of this paper.

Runs 6 and 7. The question arises, which of the two high leverage points (LTQ and AAQ) has the highest leverage? Runs 6 and 7 allow experimentation to answer this question.

Run 6 raises LTQ from 20% to 30% and leaves AAQ at 20%. The result is an optimum *false meme size* of 4.8 and 38% *percent rationalists*. Run 7 sets LTQ back to 20% and raises AAQ from 20% to 30%. The result is an optimum *false meme size* of 3.4 and 42% *percent rationalists*.

Comparing the two runs, results show that pushing on the LTQ high leverage point raised *percent rationalists* 4 percentage points. But pushing on the AAQ point raised it 8 points. AAQ has roughly twice the leverage of LTQ. This is fortunate, since AAQ training is much easier than LTQ training. Let's review why.

AAQ training (aka vote training) consists of learning two simple rules:

Rule 1. *Penalize the Deceiver* – If you discover a politician has attempted to deceive you, then when you vote or take action you should strongly oppose the politician or the source of the deception. This will have the effect of reducing attempted deception.

Rule 2. *Support the Truth Teller* – If you discover a politician has told the truth, then when you vote or take action you should strongly support the politician or the source of the truth. In this manner we encourage more truth tellers.

LTQ training in the Truth Literacy Training study consisted of learning how to spot 6 fallacies and flawed application of the Strong Evidence Rule. This required learning dozens of rules and the procedure of how to apply the Personal Truth Test, which is a much larger task than the two rules of AAQ training. Real training would require training on many more fallacies.

For brevity AAQ training was not described in the main paper. Group 1, the control group, received no training and was exposed to a neutral topic instead. Group 2 received training on how to spot deceptive claims (aka LTQ training). Group 3 received training on how to spot deceptive claims and how to take appropriate action (aka AAQ or vote training), given that knowledge. Group 3 is the trained group in the main paper.

Runs 8, 9, and 10. These simulation runs push on the two high leverage points equally to determine how much LTQ and AAQ must be raised to solve the problem.

The pattern is as LTQ and AAQ rise, more lies are detected and acted upon. This causes optimum *false meme size* to fall. Politicians are adaptive. Here they adapt by reducing the size of their lies to reduce chance of detection. They are trying to slip smaller lies through people's defenses, and it works. But it works only up to a certain point:

Run8. As LTQ and AAQ rise from 20% in run 5 to 30% in this run, optimum *false meme size* falls from 4.9 to 2. *Percent rationalists* rises from 34% to 43%.

Run 9. As LTQ and AAQ rise still further to 40%, optimum *false meme size* falls to 1.3. *Percent rationalists* rises to 49%.

Run 10. Solved mode. Finally, as LTQ and AAQ rise to 50%, optimum *false meme size* falls to 1, its lower limit. *Percent rationalists* rises to 50%, its upper limit. Raising LTQ and AAQ further has no effect, since optimum *false meme size* cannot be less than one.

Because optimum *false meme size* is now 1, the system undergoes a striking mode change once the root cause is resolved and political truth literacy changes from low to high. The tendency of politicians to tell lies does not just fall to a low level. *It disappears altogether, because now the winning strategy for politicians is telling the truth.* Those who don't tell the truth die out.

Run 10 represents the solved mode and was a counterintuitive discovery offering deep insight, a frequent occurrence when using system dynamics modeling to reveal a problem's feedback loop structure. We expected that as LTQ and AAQ rose, more degenerates would move to the Race to the Top and that loop would become dominant because it contained the most supporters.

But that's not what the model predicts will happen. *The model shows the two loops will end their perpetual duel and effectively merge into one loop, because now both loops compete for supporters by telling the truth.* The two feedback loops behave as one, because they have the same meme size of one.

Optimum *false meme size* falls from 4.9 in run 5 to 1 in run 10. This corresponds to the right moving from an extreme far-right false ideology to a truth-based moderate position, one so moderate that like the rationalists, they too pursue the common good. They are now moderates, not degenerates. There will be differences of opinion among political parties and politicians. Interpretations of what common good goals are paramount and how to achieve them will persist. But if LTQ and AAQ are high enough, citizens will elect politicians who can now work together in harmony.

Movement from partisan extremes to the moderate center eliminates support of authoritarianism. In his examination of the left and right in political systems, Bobbio (Bobbio, 1996, pp. xvi, xvii) explains why:

> …. the modern political universe is made up of two entirely separate axes: left/right and liberty /authoritarianism. …the two axes in politics combine to produce four categories: the extreme right, the moderate right, the moderate left, and the extreme left. The extremists are authoritarian, and do not accept the rules of democracy, and although the moderate left and moderate right disagree over the question of equality, they accept the same rules for the political game.

The Dueling Loops model employs a single axis, consisting of Bobbio's "the extreme right, the moderate right, the moderate left, and the extreme left." When supporters move from extreme to moderate positions, "they accept the same rules for the political game," and reject authoritarian pressure to backslide. In Bobbio's words, when the "inclusive middle" dominates, "left and right cease to be two mutually exclusive totalities like two sides of a coin

which cannot both be seen at the same time; they become two parts of a whole, a dialectic totality." (p7)

Runs 11 and 12. Some Race to the Bottom factions will adapt to rising DTQ by moderating toward the political center. But we expect others will not. They will continue to promote their deception-based far-right ideology. How would the model behave if instead of moderating, the right choses to continue a strategy of deception?

Run 11 shows what happens when instead of choosing an optimum *false meme size* of one that maximizes the number of their supporters, a group of supporters and one or more politicians feel so locked into their own false ideology that they do not adapt at all from run 5. Even though LTQ and AAQ have risen to 50%, the degenerates stick with the same level of lies used in run 5. The result is 82% rationalists, which means 18% degenerates instead of 50% degenerates.

In run 12, LTQ and AAQ are raised still further to 70%. The result is 6% degenerates. If LTQ and AAQ were raised still more to 80% (not shown), only 3% degenerates would remain.

We feel anything over about 80% is not realistic in the immediate future. Even though the Truth Literacy Training study was able to raise LTQ and AAQ to about 80% and 90%, raising political truth literacy to that high a level in a large population would probably take generations. However, we do expect that raising LTQ and AAQ to a medium level of about 50% for swing voters and the young is quite practical in the short term, in less than ten years, if a state is strongly committed to preserving democracy.

Runs 11 and 12 deal with the problem of far-right minorities who despite their small percentage of the population, have a significant effect on the political system. In a democracy the rights and desires of minorities must be respected and addressed. If 18% of voters (run 11) promote a far-right ideology, a nation will be too distracted to focus efficiently on highly demanding problems. This may be seen in the disproportionate influence far-right groups can have, such as the authoritarian populist wave in Europe of Le Pen in France, the Austrian Freedom Party in Austria, the Sweden Democrats in Sweden (who emerged from violent neo-Nazi groups in the late 1980s), and the mis-named Center Party and Forum for Democracy parties in the Netherlands. In Germany, Denney (2021) found that the far-right Alternative for Germany (AfD), even though national polls show only 10% support,

> …poses a significant and complex threat to the German constitutional order. Highly organized and openly hostile to the rules binding other political actors, the German far right has outperformed its electoral support in shaping German society. In 2020, [one of Germany's intelligence agencies] reported that the number of right-wing extremists in Germany has increased to 33,300, of whom 13,300 are thought to be willing to commit violence.

The vote share for populist parties in 32 European democracies with a lower or single house of parliament rose from 5.3% in the 1950s to 12.4% in the 2010s (Norris & Inglehart, 2019,

p. 9). Historically, the far right is where authoritarian support begins. But it cannot begin if political truth literacy is medium or high.

Run 13. The main paper stated that:

> The effect of social media false meme amplification has become large and continues to grow.[2] This amplification gives the Race to the Bottom a further advantage because of the unresolved root cause of low political truth literacy. Once the root cause is resolved, amplification no longer works.

Run 13 models this amplification by increasing *influence per degenerate* from 200 to 300. *Influence per rationalist* remains at 200. This approximates the amplification effect, since it causes the Race to the Bottom to inherently have 50% more memes than the Race to the Top. That's a huge advantage.

The result is that optimum *false meme size* is still one. *Percent rationalists* falls from 50% in run 10 to 41%. But this doesn't matter, because degenerate politicians are now telling the truth. In the real world, run 13 corresponds to a democracy that has swung to the right due to amplification, but is not dysfunctional. Healthy cooperative political debate among moderates about differences of opinion prevails, rather than dysfunctional polarization.

Experimentation shows that starting with run 13, if *false meme size* is raised from 1 to 4.9 and LTQ and AAQ are raised to 70%, the result is about the same as run 12.

Far-right behaviour

While much further research is required, the analysis and model tell us that by raising LTQ and AAQ from low to medium, the main root cause of democratic backsliding can theoretically be resolved. As that occurs, most on the right will move the center, where healthy political discourse will prevail instead of debilitating polarization. The rest of the right (the far right) will cling to their false beliefs due to the deceptive power of motivated reasoning, a well-established theory explaining how biased decision-making works (Kunda, 1990; Lodge & Taber, 2013). The theory explains why once a person is fooled into strong false political beliefs, the person becomes highly partisan and their false beliefs are unshakable.

The far-right's beliefs are so immune to truth-based arguments that they mostly cannot be changed. However, raising a nation's LTQ and AAQ reduces the ability of the far-right to recruit new members via deception. In theory, over the long term the far right will gradually diminish to such a low percentage that they will have little political impact. This is speculation, however. How various right-wing groups can best be moved to the center or eliminated altogether is an important area for further research.

This completes presentation of the simulation model. Next, we address several important concerns.

Two crucial questions

Why didn't backsliding start long ago?

The question arises: If political truth literacy (DTQ) has always been low, why didn't backsliding begin long ago? The literature contains several reasons:

1. *The shift from illegal to legal takeovers*. At first, backsliding was due to illegal takeovers. In an illegal takeover, such as military coups and invasions, DTQ matters little compared to a dictator's goals and military strength. But with the rise of the popularity of democracy, anti-pluralists switched to legal takeovers via elections characterized by deceptive "anti-pluralist rhetoric" (Luhrmann et al., 2021). This began in the 1990s, and since 2000 has accounted for about 80% of takeovers (Svolik, 2019). This helps explain why backsliding was low before the 1990s.
2. *Democracy delivered for a long time but then failed to deliver*. Per McCarthy (2019), this began in the 1970s but greatly accelerated in the 1990s, as democracy "failed to deliver the promised broad-based growth." This trend was "dramatically intensified" by the Great Recession of 2008. Prior to this, the economic benefits and personal freedom of democracy made it highly attractive. This helps explain why backsliding was low before the 1970s to 1990s period.
3. *Displacement of traditional gateway news organizations by social media*, which is far less reliable: "The rise of fake news highlights the erosion of long-standing bulwarks against misinformation in the internet age. Concern over the problem is global" (Lazer et al., 2018). "The decline in the authority of traditional social institutions began before the year 1990 and has been growing ever since" (Fukuyama, 2020). This helps explain increases in backsliding beginning in the 1990s.
4. *Injection of misinformation by international actors,* particularly Russia, into traditional news and social media. This began in the 2000s with Putin's perfection of "modern authoritarianism," now widely copied by China and others (Puddington, 2017). Russia Today TV, a combination of legitimate and propaganda news in four languages, launched in 2005 and reaches about 700 million households in over 100 countries (Staff, 2023). Russia is globally adept at social media digital propaganda. This helps explain increases in backsliding beginning in the 2000s.

Is a single high leverage point too simplistic for such a difficult problem?

A critical finding must be challenged. How can pushing on one high leverage point, *raise political truth literacy from low to high*, solve such a notoriously difficult problem? Doesn't this indicate the analysis is overly simplistic and somehow flawed? We think not for three reasons:

Reason 1. *100% of authoritarian governments depend on copious amounts of propaganda to fool their citizens into supporting them.* Evidence may be found by examination of authoritarian states. All we have examined are propaganda dependent. The pattern is so reliable that using the V-Party dataset, Luhrmann et. al. (2020) identified four key characteristics of anti-pluralism. As discussed earlier, all require political deception to implement.

100% of authoritarian leaders or parties depend on massive amounts of political deception to rise to and stay in power. Political deception works only if a population's political truth literacy is low. It follows that if it was raised to high, authoritarians would be forced to turn to another mechanism to gain and/or maintain power. What would that be? We see no plausible alternatives.

Reason 2. *The high leverage point has never been pushed on before with focused, large-scale, long-term solution elements*. Consider just Truth Literacy Training (TLT):

1. No education system has ever deeply educated students in political truth literacy, right alongside reading, writing, and math literacy.
2. No large news organization has ever made continuing education of the public in political truth literacy part of its mission, via news coverage that includes components of TLT. An example would be an article on how Hitler, Putin, and Trump each employed the classic authoritarian deception pattern of painting a false common enemy(s) 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 TV series on "The biggest lie and the most novel fallacy of the day." Or a newspaper (using manual and AI methods) might identify each deception in important political statements or speeches, state the fallacy used, and provide a link to how the fallacy works. This allows readers to continually learn how to spot patterns of deception and not be fooled. "…the ways by which we are deceived are consistent and not so hard to recognize" (Jackson & Jamieson, 2007, p. 6).
3. No major research organization has ever focused a project on developing empirically based approaches to TLT and other solution elements for pushing on the high leverage point.
4. 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 high leverage point is correct, 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 and media literacy. 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" (Ennis, 2015). 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, which Livingstone (2004) defines as "the ability to access, analyse, evaluate and create messages across a variety of contexts."

Reason 3. *Because of a complex web of intricate feedback loops, the behaviour of complex systems is highly counterintuitive.* This prevents solving difficult complex system problems without first physically modeling them, because: (Sterman, 2000, p. 27)

> The mental models people use to guide their actions are dynamically deficient. …experiments show that the greater the dynamic complexity of the environment, the worse people due relative to potential. [This is because] our cognitive maps of the causal structure of systems are vastly simplified [and] we are unable to infer correctly the dynamics of all but the simplest causal maps.

By using RCA to drive construction of a simulation model of the essential feedback loop structure of the problem, the analyst can cut through system complexity, identify well-hidden root causes and their high leverage points, and following Occam's razor, design solutions that are as simple as possible but no simpler. With the right analysis, it's possible to find ultra-high leverage points and simple solutions for pushing on them.

Proof that simple solutions can solve difficult large-scale social problems exists. Examine the Autocratic Ruler Problem in Figure 3 in the main paper. The problem existed for thousands of years and defied countless solution attempts. Being intuitively derived, these were superficial solutions doomed to failure.

Finally, after long trial and error a simple solution was tried that worked. The simple main root cause was *no easy way to replace a bad ruler with a good one*. To resolve that root cause, the solution pushed on a single ultra-high leverage point, *the concept that people have rights and therefore must have power over their rulers*. The solution itself was one of supreme simple elegance: *modern democracy, whose essence is the Voter Accountability feedback loop*. The one simple change was codification into law of the right to vote for one's rulers.

The example collection of solution elements for pushing on the single ultra-high leverage point of the backsliding problem follows this pattern. The one simple change is codification into law of the right to high political truth literacy.

Without this change the Voter Accountability feedback loop, and hence the structural design of modern democracy, is flawed and incomplete.

How political and social scientists can apply RCA using ECS diagrams

There are two processes to learn: The four-step overall problem-solving process and the nine-step process for ECS diagrams. Consider both to be a starter process to get you going. As you work, continuously improve the process as needed to fit your team and problem(s).

The overall RCA using ECS diagrams process

This is a generic process for finding and resolving a difficult large-scale social problem's root causes. Root causes are resolved by pushing on their high leverage points with solution elements. Analysis thus focuses on answering *two strategic questions*: (1) What are the high leverage points? and (2) How will the system respond to pushing on a high leverage point? The answer to the first question is expressed as ECS, which is then is used to answer the second question. Solution elements with a high probability of success are developed and recommended for implementation. Due to the magnitude of difficult large-scale social problems, implementation is managed by governments and related organizations. The process uses *four main types of experimentation*: simulation, laboratory, pilot programs, and partial implementation. The process can be summarized into four main steps:

1. *Problem formulation*: Define the problem in terms of present undesired symptoms and future desired symptoms, plus additional requirements as needed.

2. *Structural analysis*: Identify the high leverage points and their expected behaviour by identifying essential causal structure, using ECS diagrams, simulation, and other tools as necessary. A high leverage point is a solution strategy, not a solution element. One or more solution elements push on a high leverage point.

   This step answers question one and partially answers question two. This step uses the first type of experimentation, simulation, to test overall model structure and model behaviour of high leverage points.

3. *Solution convergence*: Develop solution element policy for pushing on the high leverage points. Test policy hypotheses by experimentation.

   This step fully answers question two by using more fully designed solution elements that can be tested by laboratory experiments and real-world pilot programs. This step uses the second and third types of experimentation, laboratory and pilot programs, to test solution elements. A pilot program is a simplified, small-scale, short-term form of an actual solution element.

4. *Implementation*: Implement solutions, treat the results as an experiment, and iterate until the problem is solved. While all steps are iterative, this step is especially so. Iteration allows researchers and governments (and interested third parties, such as NGOs and industry) to work together as a feedback-loop-driven learning organization, long considered a business best practice. This message was famously driven home by Peter Senge's bestselling book, *The Fifth Discipline* (1990), whose subtitle was *The Art and Practice of the Learning Organization.*

This step uses the fourth type of experimentation, partial implementation, to test the effect of solutions that are partial. For example, as described above in *Experimental confirmation*, seven solution elements were designed. Implementing only one would be partial implementation. As a second example, the first package of solution elements might push on one high leverage point. Learning from that would improve steps 2 and 3 on how to best push on a second high leverage point. A third example would be time phasing of various solution elements and the amount of each applied for various leverage points. Learning from the effects of each leverage point push would improve steps 2 and 3.

The four steps work together as a unified whole. Continuous process improvement and the feedback loop of four types of experimentation and process iteration causes solution policies to evolve over time to solve a problem to the "best" level humanly possible. This is the secret of *Kaizen*, a Japanese term meaning gradual unending process improvement. Standardize the process. Then repeat and improve it over and over, until perfection is achieved (Imai, 1986; Liker, 2004).

An iterative process makes many incremental improvements to a model until requirements are met. By contrast, a "waterfall" process is a sequential series of model construction steps with no iteration between steps. Widely used in software development, iteration is the "modern" replacement for the waterfall process (Larman & Basili, 2003). Iteration is the standard approach for large engineering design and development processes, due its ability to accommodate novelty, complexity, requirements changes, etc. (Wynn & Clarkson, 2018).

For difficult social problems iteration is required, due to the many unpredictable discoveries and the long gradual accumulation of knowledge encountered during analysis. Homer (1996), in a classic demonstration of *Why We Iterate: Scientific Modeling in Theory and Practice* using three case studies, argues that scientific modeling (where all important hypotheses are tested) is distinguished from other approaches "by the manner in which evaluation and revision are performed. … All discrepancies between model and evidence are investigated and their causes isolated to determine whether the model can not only reproduce history, but also do so for the right reasons." This cannot be done using a waterfall process, which is why scientific modelers iterate.

The ECS Diagram Construction Process

1. *Problem formulation*: Identify and precisely define a causal problem with repeated solution failures, or solutions that achieved far less that you suspect is possible.
2. *List solutions that failed*: If there are many, organize then into categories.
3. *Start on the superficial layer*: Using pencil and paper at first, quickly sketch the superficial layer of a ECS diagram. Start at problem symptoms and work backward from there, by asking "WHY did this occur?" For each intermediate cause, do as described in examples A and B in the main paper by asking "Is this an intermediate or root cause?" This will allow building out the superficial layer using the list of solutions that failed.

4. *Iterate on the superficial layer*: Assume your first version will be weak. Iterate until the diagram feels strong and is providing deep insights you lacked before.
5. *Stop superficial layer iteration*: Stop when it's hard to go beyond the lowest intermediate cause. This indicates you are probably encountering the fundamental layer and need to begin building a feedback loop simulation model.
6. *Start on the fundamental layer*: Start with a causal loop diagram, a standard best practice. Build it by continuing to use RCA. Simultaneously, attempt to complete the fundamental layer of the diagram. Update the superficial layer as necessary as your iterations proceed.
7. *Switch from causal loop diagramming to feedback loop modeling*: When you've gone as far as you can go with step 6, convert the causal loop diagram into a system dynamics simulation model and proceed until the four requirements for a comprehensive theory are met. Other types of simulation may be used. However, we recommend system dynamics for first simulation work due to its focus on feedback loop structure.
8. *Test key assumptions as you go*: As the analysis begins to mature by becoming more and more able to explain the four forces of ECS diagrams, test all key assumptions as you go. This is particularly important for the simulation model, though it applies to all process steps.
9. *Stop iteration when the analysis goal is achieved*: The standard goal is the four requirements for a comprehensive theory of a difficult large-scale social problem. Custom requirements may be added as part of problem formulation, such as solving the problem by a certain deadline or keeping solution success probability above 95%.

To gain familiarity with thinking in terms of ECS diagrams, we recommend that you first perform retrospective analyses of solved problems. Start by diagraming the chronic knee pain problem described in Definitions in the main paper. Then diagram examples A and B in Figure 3, main paper. After that, choose problems you understand well. These exercises will be much easier because what solutions worked is known. After 10 to 20 practice problems, switch to unsolved problems.

These educational resources should be helpful:

1. For an introduction to RCA see the book *Root Cause Analysis: The Core of Problem Solving and Corrective Action*, Duke Okes, 2019.
2. This online article[3] briefly explains ECS diagrams (also called social force diagrams), using nine examples. These were created using Microsoft Visio, a diagramming tool.
3. For a state-of-the-art review of step 1 in the process above and an RCA-based problem-solving process based on Toyota's A3 method, see the paper on *The Most Underrated Skill in Management*, by Repenning, Kieffer, and Astor, 2017, Sloan

Management Review. The authors, who "have studied and worked with dozens of organizations and taught over 1,000 executives," found that "problem formulation is the single most underrated skill in all of management practice."

The A3 method differs greatly from ECS diagrams. A3 is a general-purpose RCA-based method for typical business problems of a wide variety. ECS diagrams are designed for a more demanding and narrower class of problems: difficult large-scale social problems. Still, review of the paper should provide some insights into how to apply and manage a project using RCA using ECS diagrams.

4. For an advanced in-depth description of the world's best organization-wide application of RCA, see *The Toyota Way: 14 Management Principles from the World's Greatest Manufacturer*, Jeffrey Liker, 2004 (the first edition, not the second edition of 2021). Be sure to study Figure 1-2, page 13, and Figure 20-3, page 256. These diagrams summarize the culture and process.

   Think of auto manufacturing as one of many industries that has mastered RCA. An industry is the equivalent of one of many fields of social science that, we expect, will also eventually master RCA, *using the principles expounded in this book* as a starting point. The book describes a manufacturing process. *What is it that social scientists manufacture, as they strive to solve difficult problems? It is a long steady stream of incremental improvements and occasional system engineering breakthroughs, such as resolving main root causes.* From this perspective, what has been learned by Toyota and other companies can be applied to social problems.

5. The standard textbook for learning causal loop diagrams and system dynamics is *Business Dynamics: Systems Thinking and Modeling for a Complex World*, John Sterman, 2000.

Notes

[1] We originally used a longer incubation time of 5 years. However, this caused the first five years of some of the model graphs to have distracting straight lines. Because of this we changed to 1 year. This had no effect on the model after 5 years.

[2] For example, see Fisher's (2022, pp. 5, 9) description of how Facebook's recommendation engine amplifies false memes and how large the effect has become. "The more incendiary the post, the more widely the platforms spread it. … Our algorithms exploit the human brain's attraction to divisiveness."

[3] See https://www.thwink.org/sustain/glossary/SocialForceDiagrams.htm.

References

Bobbio, N. (1996). *Left and right: The significance of a political distinction*. The University of Chicago Press.

Dawkins, R. (1976). *The Selfish Gene*. Oxford University Press.

Denney, S. (2021). The German Far Right Doesn't Need to Win Elections to Be Dangerous. *Lawfare*. https://www.lawfareblog.com/german-far-right-doesnt-need-win-elections-be-dangerous

Dillard, J., & Shen, L. (2013). *The SAGE Handbook of Persuasion: Developments in Theory and Practice*. SAGE.

Ennis, R. (2015). Critical Thinking: A Streamlined Companion. In *The Palgrave Handbook of Critical Thinking in Higher Education* (pp. 31–47). Palgrave Macmillan.

Fisher, M. (2022). *The Chaos Machine: The Inside Story of How Social Media Rewired Our Minds and Our World*. Little, Brown and Company.

Forrester, J. (1971). *World Dynamics*. Wright-Allen Press.

Fukuyama, F. (2020). 30 Years of World Politics: What Has Changed? *Journal of Democracy*, *31*(1), 11–21. https://doi.org/10.1353/jod.2020.0001

Ghaffarzadegan, N., Lyneis, J., & Richardson, G. P. (2011). How small system dynamics models can help the public policy process: N. Ghaffarzadegan *et al.* : Small System Dynamics Models. *System Dynamics Review*, *27*(1), 22–44. https://doi.org/10.1002/sdr.442

Homer, J. B. (1996). Why we iterate: Scientific modeling in theory and practice. *System Dynamics Review*, *12*(1), 1–19. https://doi.org/10.1002/(SICI)1099-1727(199621)12:1<1::AID-SDR93>3.0.CO;2-P

Imai, M. (1986). *Kaizen: The Key to Japan's Competitive Success*. McGraw Hill.

Jackson, B., & Jamieson, K. H. (2007). *Unspun: Finding Facts in World of Disinformation*. Random House.

Kunda, Z. (1990). The Case for Motivated Reasoning. *Psychological Bulletin*, *108*, 480–498.

Larman, C., & Basili, V. (2003). Iterative and Incremental Development: A Brief History. *Computer*, *36*(6). https://www.it.uu.se/edu/course/homepage/acsd/vt08/SE1.pdf

Lazer, D., Baum, M., Benkler, Y., Berinsky, A., Greenhill, K., Menczer, F., Metzget, M., Nyhan, B., & Pennycook, G. (2018). The science of fake news: Addressing fake news requires a multidisciplinary effort. *Science*.

Liker, J. (2004). *The Toyota Way: 14 Management Principles from the World's Greatest Manufacturer*. McGraw Hill.

Livingstone, S. (2004). What Is Media Literacy? *Intermedia*, *32*(2). https://www.aspeninstitute.org/wp-content/uploads/2010/11/Digital_and_Media_Literacy.pdf

Lodge, M., & Taber, C. (2013). *The Rationalizing Voter*. Cambridge University Press.

Luhrmann, A., Medzihorsky, J., Hindle, G., & Lindbert, S. (2020). *New Global Data on Political Parties: V-Party*. V-Dem Institute. chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.v-dem.net/static/website/img/refs/vparty_briefing.pdf

Luhrmann, A., Medzihorsky, J., & Lindberg, S. I. (2021). Walking the Talk: How to Identify Anti-Pluralist Parties. *Working Paper*. https://gupea.ub.gu.se/bitstream/handle/2077/68137/gupea_2077_68137_1.pdf

McCarthy, J. (2019). Authoritarianism, Populism, and the Environment: Comparative Experiences, Insights, and Perspectives. *Annals of the American Association of Geographers*, *109*(2), 301–313. https://doi.org/10.1080/24694452.2018.1554393

Meadows, D., Behrens, W., Meadows, D., Randers, J., & Zahn, E. (1974). *Dynamics of Growth in a Finite World*. Wright-Allen Press.

Meadows, D., Meadows, D., Randers, J., & Behrens, W. (1972). *The Limits to Growth*. Potomac Associates.

Noel, A., & Therueb, J.-P. (2008). *Left and Right in Global Politics*. Cambridge University Press.

Norris, P., & Inglehart, R. (2019). *Cultural Backlash: Trump, Brexit, and the Rise of Authoritarian Populism*. Cambridge University Press.

Puddington, A. (2017). *Breaking Down Democracy: Goals, Strategies, and Methods of Modern Authoritarians* (p. 64). FreedomHouse.org.

Senge, P. (1990). *The Fifth Discipline: The Art and Practice of the Learning Organization*. Currency Doubleday.

Staff. (2023). *Distribution of Russia Today*. Russia Today. https://www.rt.com/about-us/distribution/

Sterman, J. (2000). *Business Dynamics: Systems Thinking and Modeling for a Complex World*. Irwin McGraw-Hill.

Svolik, M. W. (2019). Polarization versus Democracy. *Journal of Democracy*, *30*(3), 20–32. https://doi.org/10.1353/jod.2019.0039

Sze To, G. N., & Chao, C. Y. H. (2010). Review and comparison between the Wells–Riley and dose-response approaches to risk assessment of infectious respiratory diseases. *Indoor Air*, *20*, 2–16. https://doi.org/10.1111/j.1600-0668.2009.00621.x

Voters Rarely Switch Parties, but Recent Shifts Further Educational, Racial Divergence. (2020). PEW Research Center. https://www.pewresearch.org/politics/2020/08/04/voters-rarely-switch-parties-but-recent-shifts-further-educational-racial-divergence/

Wynn, D., & Clarkson, P. J. (2018). Process models in design and development. *Research in Engineering and Design*, *29*, 161–202.