3-2022

Stress Testing Governance

Rory Van Loo
Boston University School of Law

Stress Testing Governance

Rory Van Loo[*]

In their efforts to guard against the world's greatest threats, administrative agencies and businesses have in recent years increasingly used stress tests. Stress tests simulate doomsday scenarios to ensure that the organization is prepared to respond. For example, agencies role-played a deadly pandemic spreading from China to the United States the year before COVID-19, acted out responses to a hypothetical hurricane striking New Orleans months before Hurricane Katrina devastated the city, and required banks to model their ability to withstand a recession prior to the economic downturn of 2020. But too often these exercises have failed to significantly improve readiness for the subsequent crises. This Article shows that stress tests are used more widely than is commonly assumed, reaching well beyond financial regulation. It then argues that administrative stress tests should be seen as potentially powerful tools for administrative governance, but ones that suffer from significant shortcomings as currently deployed. Most notably, stress tests lack adequate transparency, oversight, and imagination. Also, they are too often voluntary for businesses and agencies whose performance failures could have great societal ramifications. By depriving stakeholders of crucial information about organizational readiness, these shortcomings weaken the nation's ability to prevent and prepare for disasters. Preparing for disasters will only become more important as technologies transform everything from stock trading to elections and climate change creates more volatile weather. With improved design and wider deployment, stress tests have the potential to become a central tool for public and private accountability in an era of escalating societal risks.

* Associate Professor of Law, Boston University; Affiliated Fellow, Yale Law School Information Society Project. I am grateful to Chris Havasy, Gary Lawson, Mira Marshall, Nicholas Parrillo, Christopher Robertson, David Walker, Robert Weber, and David Zaring, and participants at the National Business Law Scholars Conference for extremely helpful comments. Victoria Abramchuk, Brianne Allan, Jennifer Bentley, Celene Chen, Leah Dowd, Derek Farquhar, Christopher Hamilton, Ian Horton, Ryan Kramer, Jack Langa, Christina Luo, Tyler Stites, and Brittany Swift contributed excellent research. I appreciate the Vanderbilt Law Review Editors' sharp eyes, patience, and helpfulness throughout the editing process. First submitted for publication March 22, 2020.

INTRODUCTION

Stress tests emerged in the sixteenth century as an engineering exercise for determining whether a structure might break when faced with an external shock such as a storm, an earthquake, or flooding.[1] They serve to focus engineers' minds on worst-case scenarios—such as an airplane wing or bridge breaking—that people tend to ignore because they subconsciously avoid thinking about unlikely, terrifying events.[2] By forcing the engineer to simulate that worst-case scenario, stress tests also aim to provide information about structural changes that might be needed to prevent catastrophic collapses.[3]

Centuries after their emergence in engineering, the administrative state has begun to broadly embrace stress tests as a way

1. *See* ARNOLD W. HENDRY, ELEMENTS OF EXPERIMENTAL STRESS ANALYSIS 63 (1964).
2. *See id.*
3. *See id.*

to prepare for dire circumstances.[4] The Pentagon regularly conducts war games, including preparing for a North Korean nuclear attack through a computer simulation of command headquarters and over thirty thousand troops on the ground in South Korea participating in corresponding drills.[5] The Federal Aviation Administration issued a rule stating that all major airports at least once every three years "must hold a full-scale airport emergency plan exercise," including for bombs, fires, natural disasters, and hijackings.[6] And after a tsunami triggered the meltdown of nuclear plant reactors in Fukushima, Japan, which contaminated nearby communities with radiation, regulators around the globe obliged power plant operators to simulate related breakdowns.[7]

Each of these contexts shows a government agency using stress tests broadly defined,[8] in that a simulation aims to assess critical breaking points by subjecting the organization to extreme conditions. Stress tests are targeted at threats that allow too little time to act to prevent large-scale harm once they materialize.

Despite having the potential to make a difference between life and death on a mass scale, the widespread use of agency stress tests has gone unidentified in the legal literature. When scholars have turned their attention to stress tests, the focus has mostly been on regulators' stress testing of financial institutions, rather than stress testing administrative agencies or other businesses.[9]

---

4. Legal scholars mostly use the term stress test in the context of finance. This Article's broad view of stress tests reflects the widespread use of the term in engineering, medicine, the military, and elsewhere. *See, e.g.*, Peter Behr, *Multinational Panel Calls for Tougher 'Stress Tests' of Nuclear Plant Safety Systems*, N.Y. TIMES (June 2, 2011), https://archive.nytimes.com/www.nytimes.com/cwire/2011/06/02/02climatewire-multinational-panel-calls-for-tougher-stress-33984.html?pagewanted=2 [https://perma.cc/8XWR-K7HN] (referring to nuclear plant exercises as stress tests).

5. The Editorial Board, Opinion, *Trump Gets Exercised over Exercises,* WALL ST. J. (Mar. 6, 2019, 7:18 PM), https://www.wsj.com/articles/trump-gets-exercised-over-exercises-11551917921 [https://perma.cc/7VGQ-NJX9] ("[S]tress-testing U.S. forces is not a waste of money.").

6. 14 C.F.R. § 139.325(h) (2021).

7. *See generally* Stephen G. Burns, *The Fukushima Daiichi Accident: The International Community Responds*, 11 WASH. U. GLOB. STUD. L. REV. 739 (2012) (discussing the international response, including using stress tests to determine whether other plants were prepared).

8. This Article's contributions do not depend on any particular definition of stress tests. In the legal literature the term has typically been used more narrowly to refer mostly to bank stress tests, but it has broader use in other academic fields and among public leaders. *See, e.g.*, Kayla Tausche, *White House Studying Supply Chain 'Stress Tests' After Semiconductor Shortages, Sources Say*, CNBC, https://www.cnbc.com/2021/04/01/white-house-studying-supply-chain-stress-tests-after-semiconductor-shortages-sources-say-.html (last updated Apr. 1, 2021, 12:21 PM) [https://perma.cc/Q324-ZJBB] ("As part of an ongoing review into critical supply chains, the Biden administration is considering requesting that supply chains undergo 'stress tests' of hypothetical scenarios . . . .").

9. *See, e.g.*, Mehrsa Baradaran, *Regulation by Hypothetical*, 67 VAND. L. REV. 1247, 1318 (2014) (proposing improvements to financial regulators' use of hypotheticals); Kathryn Judge,

This Article marks stress tests as central tools of modern governance. It demonstrates the breadth of the administrative state's use of stress tests across diverse agencies. It then diagnoses why agencies too often ignore the alarm bells that the tests sound, like when the Department of Health and Human Services in 2019 surfaced critical shortcomings after getting hundreds of government officials to role-play their responses to a pandemic that would kill 110 million people worldwide.[10] Finally, the Article sketches an expanded vision for stress tests to combat the rising complexity of crises in an age of economic insecurity, digital innovation, and global warming.

Despite scholarly inattention, stress tests intersect with vibrant administrative law conversations. The dominant preoccupation of administrative law is the accountability of unelected bureaucrats.[11] In that literature, accountability advances both legitimacy and effectiveness.[12] A significant strand of administrative law scholarship

---

Stress Testing During Times of War 1 (Eur. Corp. Governance Inst., Working Paper No. 529/2020, 2020), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3633310 [https://perma.cc/45M8-PLGY] (arguing that bank "stress tests become more important, not less, in the midst of systemic distress, but only if the stress scenarios are modified to reflect the distinct challenges an economy is facing"); Robert Weber, *A Theory for Deliberation-Oriented Stress Testing Regulation*, 98 MINN. L. REV. 2236, 2243 (2014) (offering reforms for financial stress tests). To elaborate, although those articles provide valuable insights on which this Article builds, their focus is mostly different in two main ways. Most importantly, because the financial regulation stress test literature is about a particular industry, it does not examine the breadth of stress tests across the administrative state. Additionally, bank stress tests are not about stress testing the regulators—they are about stress testing the banks. This Article, in contrast, focuses on stress testing agencies. This Article also builds on two innovative articles focused on bank regulation that begin to go in the direction of considering agencies, but in narrower ways. First, John Crawford proposes training exercises for bank regulators to address inexperience. John Crawford, *Wargaming Financial Crises: The Problem of (In)experience and Regulator Expertise*, 34 REV. BANKING & FIN. L. 111, 113 (2015). More recently, Matthew Turk proposes using bank stress tests to analyze the rules that financial regulators have written for stress tests. Matthew C. Turk, *Stress Testing the Banking Agencies*, 105 IOWA L. REV. 1701, 1750–51 (2020). Besides differing from this Article by focusing on financial regulation, Turk's proposal is still not about stress testing agencies—despite the title—but instead on testing the rules they write for bank stress tests. Both of those valuable financial regulation articles differ from this Article's emphasis on stress tests created for the purpose of testing the agencies' overall preparedness for a crisis on an organizational (or administrative) level.

10. *See infra* Section I.B.

11. *See, e.g.*, Margaret H. Lemos, *Democratic Enforcement? Accountability and Independence for the Litigation State*, 102 CORNELL L. REV. 929, 942 (2017) (explaining accountability).

12. *See* Lisa Schultz Bressman, *Disciplining Delegation After* Whitman v. American Trucking Ass'ns, 87 CORNELL L. REV. 452, 485 (2002) ("[Administrative law principles] require agencies in general to articulate a basis for their policy determinations and, in particular, to articulate the standards for those determinations."); Jessica Mantel, *Procedural Safeguards for Agency Guidance: A Source of Legitimacy for the Administrative State*, 61 ADMIN. L. REV. 343, 348 (2009) ("[E]nsuring that our government meets the public's needs depends on not only promoting the legitimacy of the administrative state, but also preserving the administrative state's ability to effectively exercise its powers."); Edward Rubin, *The Myth of Accountability and the Anti-Administrative Impulse*, 103 MICH. L. REV. 2073 (2005) (embracing more expansive concept of accountability).

can thus be seen as examining tools that provide better oversight of how well government actors are deploying the authority delegated to them. Congress arguably has a "constitutional duty to supervise" administrative agencies.[13] Moreover, because administrative law has traditionally focused on rulemaking and formal adjudication, scholars have called for the field to better reflect the full array of activities that agencies undertake.[14]

This Article responds to those calls by showing how well-designed stress tests can provide Congress with a mechanism to supervise agencies' readiness to safeguard society. Furthermore, the modern era of regulation and risk management is characterized by heavy privatization in that businesses often perform the key risk management functions that the state might otherwise assume.[15] Thus, mandating stress testing of banks, oil companies, or other businesses is related to mandating stress tests of administrative agencies in the sense that both private firms and public agencies are entities along the chain of delegated public authority.[16]

Another strand of the literature has analyzed the broader sphere of activity on which stress tests arguably build: risk management. That research shows how agencies quantitatively model the risks of the next financial, environmental, or public health calamity.[17] Sophisticated risk

---

13. *See* Gillian B. Metzger, *The Constitutional Duty to Supervise*, 124 YALE L.J. 1836, 1874–1904 (2015).

14. *See* Edward Rubin, *It's Time to Make the Administrative Procedure Act Administrative*, 89 CORNELL L. REV. 95, 96–97 (2003) (highlighting many ways in which the APA fails to reflect large parts of what administrative agencies actually do and concluding "[i]t is time to rewrite the Administrative Procedure Act"); Lemos, *supra* note 11 (emphasizing unlocking the black box of agency enforcement).

15. *See, e.g.*, Cary Coglianese & David Lazer, *Management-Based Regulation: Prescribing Private Management to Achieve Public Goals*, 37 LAW & SOC'Y REV. 691, 691, 726 (2003); Jody Freeman, *The Private Role in Public Governance*, 75 N.Y.U. L. REV. 543, 549–56 (2000); Gillian E. Metzger, *Privatization as Delegation*, 103 COLUM. L. REV. 1367, 1369 (2003); Martha Minow, *Public and Private Partnerships: Accounting for the New Religion*, 116 HARV. L. REV. 1229, 1237–42 (2003).

16. *See, e.g.*, Kenneth A. Bamberger, *Regulation as Delegation: Private Firms, Decisionmaking, and Accountability in the Administrative State*, 56 DUKE L.J. 377, 384 & nn.12–13 (2006) (discussing privatized risk management as delegation).

17. For a review of risk-based regulation, or risk management, see generally, for example, *id.* at 467–68 (discussing "[i]ncorporating organizational learning approaches into regulatory and agency design" as a means of managing risk); Cary Coglianese & Gary E. Marchant, *Shifting Sands: The Limits of Science in Setting Risk Standards*, 152 U. PA. L. REV. 1255 (2004) (evaluating the EPA's approach to risk management); Julie E. Cohen, *The Regulatory State in the Information Age*, 17 THEORETICAL INQUIRIES L. 369, 390 (2016) (underscoring the growing importance of regulators managing risks); James Fanto, *Anticipating the Unthinkable: The Adequacy of Risk Management in Finance and Environmental Studies*, 44 WAKE FOREST L. REV. 731 (2009) (covering risk management in financial and environmental regulation); Kathryn Judge, *Fragmentation Nodes: A Study in Financial Innovation, Complexity, and Systemic Risk*, 64 STAN. L. REV. 657, 659–61 (2012) (depicting inadequate financial risk management efforts); and J.B.

modeling, however, does not mean that agencies already undertake stress tests. Stress tests go beyond modeling threats, such as the risk of a natural disaster or financial crisis, to assess how well private or public organizations would respond to those threats. Stress tests thus incorporate elements of scenario analysis and simulations.[18]

To illustrate how stress tests fit into the broader toolkit for risk management, consider by analogy how cardiologists regularly conduct "stress tests" by putting patients on treadmills and gradually ramping up the speed to see how their bodies react when pushed.[19] Poor results prompt preventive advice to exercise more, eat better, or take medication.[20] In contrast, agencies' risk projections analogize to a different procedure undertaken by cardiologists: the sophisticated analysis of bloodwork, which can be used to give the patient health information. Although valuable, analyzing blood is not the same as conducting a treadmill test.[21] Similarly, agency risk analytics do not test *organizational* readiness in the same way as would a stress test requiring agency leaders to make decisions, develop plans, and solve problems in the simulation of, say, a nuclear attack.

The difference is subtle, but the stakes are high. Medical stress tests have saved countless lives by alerting people to heightened risks of heart disease—the leading cause of death worldwide.[22] Evidence

---

Ruhl, *Managing Systemic Risk in Legal Systems*, 89 IND. L.J. 559, 602–03 (2014) (applying complexity science to legal regimes for environmental and other areas of disaster management).

18. The boundary lines between risk management tools are often blurred. Stress tests have significant characteristics that overlap with various scenario planning tools and could be broadly classified as a type of scenario planning—indeed, robust stress tests incorporate scenario analyses. *See infra* Section II.C. Stress tests could be seen as a more organizationally focused and simulation-intensive form of scenario planning. *Cf.* Axel Volkery & Teresa Ribeiro, *Scenario Planning in Public Policy: Understanding Use, Impacts and the Role of Institutional Context Factors*, 76 TECH. FORECASTING & SOC. CHANGE 1198, 1199 (2009) (adopting a broad definition of scenario planning as reaching threats, policies, and organizational blind spots). However, stress tests also combine elements of war games, or simulations, which are sometimes viewed as distinct from scenario planning. *See, e.g.*, Jan Oliver Schwarz, Camelia Ram & René Rohrbeck, *Combining Scenario Planning and Business Wargaming to Better Anticipate Future Competitive Dynamics,* 105 FUTURES 133, 135 (2019) (distinguishing wargaming from scenario planning). How to classify stress tests is less important than determining how to optimize stress tests' contributions to governance. The broad literature on scenario planning provides helpful foundations but scenario planning has so many different forms that the literature tends to pay insufficient attention to key stress test characteristics of simulative testing and organizational readiness, which are particularly important in the context of administrative accountability. *See, e.g.*, Volkery & Ribeiro, *supra* (reviewing the literature on scenario planning with limited attention to these characteristics).

19. *See generally* Stephen N. Morris & Paul L. McHenry, *Role of Exercise Stress Testing in Healthy Subjects and Patients with Coronary Heart Disease*, 42 AM. J. CARDIOLOGY 659 (1978).

20. *Id.* at 663.

21. *See infra* Part I (describing stress tests used by administrative agencies and how they differ from risk management practices).

22. Morris & McHenry, *supra* note 19, at 659.

suggests that in at least some organizational contexts—such as firefighting units, air traffic control towers, and other high-reliability organizations—intense simulations can meaningfully increase the chances of success in the face of an emergency.[23] Yet when designed poorly, they can have little or no effect, thereby simply wasting resources better used elsewhere.[24] Effective stress tests thus can mean the difference between success and failure on a large scale.

Moreover, stress tests are tools well suited to the modern era. Administering the law has become far more complicated because technologies enable catastrophes on a previously unimaginable scale.[25] Weather shocks to our infrastructure will become more frequent and intense due to climate change.[26] Stress tests are the kind of tool that may be crucial to reach the next level of administrative risk management for an increasingly complex world.

It is worth considering whether a new paradigm is needed for stress tests to fulfill their governance potential. Agencies need regular assessments of their organizational readiness for larger scale threats. But left to their own devices, they often design stress tests in ways that obscure institutional failures, which saves face but makes learning—and valuable change—less likely.[27] To maximize stress test benefits, in at least some important contexts the law should involve actors outside

23. *See, e.g.*, KARL E. WEICK & KATHLEEN M. SUTCLIFFE, MANAGING THE UNEXPECTED: RESILIENT PERFORMANCE IN AN AGE OF UNCERTAINTY 1–2 (2d ed. 2007) (referring to diverse organizational responses to emergencies); Jos A. Rijpma, *From Deadlock to Dead End: The Normal Accidents—High Reliability Debate Revisited*, 11 J. CONTINGENCIES & CRISIS MGMT. 37, 37–39 (2003) (summarizing the academic literature on high-reliability organizations); *see also* Weber, *supra* note 9, at 2243 (applying "high-reliability organizations" to financial regulation). The evidence is strongest for individuals participating in emergency simulations. *See, e.g.*, Alexis Borg Sapiano, Roberta Sammut & Josef Trapani, *The Effectiveness of Virtual Simulation in Improving Student Nurses' Knowledge and Performance During Patient Deterioration: A Pre and Post Test Design*, 62 NURSE EDUC. TODAY 128 (2018) (finding that students who took part in virtual simulations testing complex emergency situations, such as rapid patient deterioration, tested better after performing in the scenario); Robert T. Hays, John W. Jacobs, Carolyn Prince & Eduardo Salas, *Flight Simulator Training Effectiveness: A Meta-Analysis*, 4 MIL. PSYCH. 63 (1992) (examining twenty-six experiments testing the efficacy of flight simulator training and noting which characteristics provided the most benefits); Ioana Koglbauer, *Simulator Training Improves Pilots' Procedural Memory and Generalization of Behavior in Critical Flight Situations*, 20 COGNITION BRAIN BEHAV. 357 (2016) (finding that pilots who underwent simulated training performed better in critical flight situations).

24. *See infra* Section I.C.

25. David L. Alderson & John C. Doyle, *Contrasting Views of Complexity and Their Implications for Network-Centric Infrastructures*, 40 IEEE TRANSACTIONS ON SYS. MAN & CYBERNETICS 839, 843 (2010) ("[M]odern institutions and technologies . . . enable catastrophes on a scale unimaginable without them . . . .").

26. *See, e.g.*, A. Dan Tarlock & Deborah M. Chizewer, *Living with Water in a Climate-Changed World: Will Federal Flood Policy Sink or Swim?*, 46 ENV'T L. 491, 491–92 (2016) (summarizing the consensus on climate change implications).

27. *See infra* Part I.

the responsible agency in the design, management, and evaluation of stress tests. Those actors would ideally also help address stress test deficits in imagining future threats and conveying the full consequences of failing to prepare—including the often underappreciated economic costs of crises.

Involvement by the public, Congress, or the White House would help. A new risk management group, whether standalone or located in an existing governmental entity such as the Government Accountability Office, could also play a meta-agency stress test supervisory role. The new paradigm needed is thus to shift from viewing stress tests as voluntary agency exercises, or something mandated of banks, toward viewing them as broader legal tools for governance accountability.

The challenging task would then be to determine when the costs justify the accountability benefits. At a minimum, the significant current resources devoted to stress tests throughout the administrative state could be put to better use. If infused by the law with external oversight, stress tests will have greater potential to jostle a too often inert administrative state into greater efficacy in safeguarding society.

Part I shows the widespread use of stress tests and their shortcomings by looking at three case studies in public health, the environment, and the financial system. Part II begins the task of reconceptualizing these exercises as big-picture, future-oriented, and transparent tests of the government. Part III illustrates how revived stress tests could improve agencies' ability to guard against pandemics, environmental disasters, and financial crises. It also briefly considers contexts, such as supply-chain management and cybersecurity, in which mandates might be extended to businesses. Stress tests have the potential to play a considerably broader role than they currently do in both public- and private-sector accountability.

Before turning to the main discussion, two notes on scope and limits are in order. This Article integrates discussions of stress testing of administrative agencies with stress testing of private firms. Although these two contexts differ in meaningful ways, they also have important similarities given the goal here of framing stress tests as a governance tool. The meaningful differences between these contexts are worthy of further study and elaboration. Moreover, not all of the discussion below, which focuses more on stress testing administrative agencies, applies to mandated stress testing of firms. Nonetheless, many of this Article's core high-level themes, such as stress test shortcomings and potential, apply to both contexts.

Also, the direct empirical basis for expanding administrative stress tests is limited. That limitation reflects the broader difficulty in

studying rare events, which are by definition so small in number as to defy robust statistical study. Moreover, each oil spill or financial crisis tends to unfold differently, with a complex variety of politics, funding constraints, and other factors contributing to any given decision not to take a preparatory step.[28] Consequently, it is almost impossible to know either why leaders failed to take adequate precautions or the counterfactual of what would have happened if stress tests had been used more effectively. It is also difficult to quantify the benefits of crisis prevention, or of improving accountability—as is the case with other pervasive administrative tools, such as notice-and-comment rulemaking. This Article's recommendations should thus be viewed as working hypotheses identifying sensible steps based on an inference from what is known about simulations, the limitations of organizational preparedness, and accountability. The main aim here is to draw attention to these tools' widespread use, to call for greater study of them, and to theorize their broader potential role in governance.

Despite epistemic limitations, the widespread adoption of stress tests in construction and cardiology was driven not by statistically significant evidence that they worked, but by common sense innovation.[29] Lawmakers and agency leaders are the primary architects and caretakers of the administrative state. Just as no architect would construct a major building without considering what can be learned from stress tests, legal designers should give greater attention to integrating stress tests into the continually updated blueprints of the administrative state.

I. THE PROMISE, PERVASIVENESS, AND FAILURES OF STRESS TESTS

Administrative stress tests mostly aim to improve the organizational management of crises. They may do so by enhancing an agency's ability to prevent the emergency from materializing, or by promoting better advanced planning so that the agency can reduce the harm from the adverse event once it occurs. To provide institutional context for how stress tests might improve those difficult tasks, this Part considers the state's lack of preparation, underlying psychological explanations, and current use of stress tests.

---

28. *See infra* Section II.C.

29. *See* Mohamed K. ElBatanouny, Gregor Schacht & Guido Bolle, *History of Load Testing of Bridges*, *in* 12 LOAD TESTING OF BRIDGES 9, 11 (Eva O.L. Lantsoght ed., 2019) ("[I]nitially no empirical data or previous knowledge was available with respect to what constituted allowable or critical structural states."); Michael W. Luong, Maya Ignaszewski & C.M. Taylor, *Stress Testing: A Contribution from Dr. Robert A. Bruce, Father of Exercise Cardiology*, 58 BC MED. J. 70, 71 (2016) (describing how Dr. Robert Bruce revolutionized exercise stress tests in cardiology).

A. The Organizational Aims of Stress Tests

1. Addressing the Psychological Limits of Crisis Management

At their engineering core, stress tests produce information that is valuable for evaluating readiness for hypothetical scenarios and responding to the human mind's psychological limits.[30] Although the information-gathering goal of stress tests is straightforward, the psychological dimension would benefit from brief elaboration.

When imagining the likelihood of future events occurring, people envision familiar scenarios—a phenomenon known as the "availability heuristic."[31] The mind tends not to think about rare events, especially ones that the individual has never experienced.[32] Thus, one challenge for crisis management is getting people to think about something that the mind tends to avoid.

Nor do near misses necessarily help to prompt action. A case study of the underlying psychology comes from space travel.[33] In several launches, NASA scientists observed a threat: insulation foam shedding from an external fuel tank, which once even damaged a wing in the process.[34] Instead of learning from the evidence that a dangerous crash risk existed, however, they interpreted the absence of a crash as a success—an "outcome bias."[35] This encouraged the inference that the foam shedding was not as dangerous as originally thought, or at least not as worthy of attention.[36] That bias proved fatal to seven crewmembers in 2003, when foam shedding punctured a hole in the wing of the Columbia space shuttle, causing the shuttle to disintegrate upon reentry into the atmosphere.[37] Counterintuitively, in the wake of

30. *See* HENDRY, *supra* note 1; Steven L. Schwarcz, *Regulating Derivatives: A Fundamental Rethinking*, 70 DUKE L.J. 545, 585 (2020) (discussing stress tests as a means of addressing cognitive biases).

31. Amos Tversky & Daniel Kahneman, *Availability: A Heuristic for Judging Frequency and Probability*, 5 COGNITIVE PSYCH. 207, 229 (1973) ("In thinking of [rare] events we often construct *scenarios* . . . . The plausibility of the scenarios that come to mind, or the difficulty of producing them, then serve as a clue to the likelihood of the event.").

32. *Id.*

33. Robin L. Dillon & Catherine H. Tinsley, *How Near-Misses Influence Decision Making Under Risk: A Missed Opportunity for Learning*, 54 MGMT. SCI. 1425, 1427–28 (2008).

34. 1 NAT'L AERONAUTICS & SPACE ADMIN., REPORT OF COLUMBIA ACCIDENT INVESTIGATION BOARD 128 fig.6.1-7 (2003).

35. *See* 1 *id.*; Jonathan Baron & John C. Hershey, *Outcome Bias in Decision Evaluation*, 54 J. PERSONALITY & SOC. PSYCH. 569, 569 (1988).

36. Dillon & Tinsley, *supra* note 33, at 1427–28.

37. 1 NAT'L AERONAUTICS & SPACE ADMIN., *supra* note 34, at 121.

close calls *without any consequences*, people across diverse organizations tend to lower their estimates of accompanying risks.[38]

The human mind fares better in the wake of actual tragedies. Once a tragedy occurs, people tend to invest in additional precautions in the *immediate* aftermath.[39] However, since rare crises that occurred many years ago will not be "available" to the mind,[40] over time they may return to being overlooked or discounted.[41] These psychological challenges, exacerbated by organizational dynamics and market failures,[42] help explain why policymakers have added stress tests to their toolkits. A stress test seeks to address these psychological shortcomings by causing people to regularly experience a simulated crisis, rather than a real-world near miss, thereby keeping the threat and its consequences top of mind.

2. Improving Existing Administrative Risk Management

Stress tests layer onto a larger administrative state apparatus for managing risks. The risk management toolkit includes threat forecasting, scenario analysis, and other tools that are designed to unearth the likelihood and nature of future challenges such as natural disasters.[43] The law often imposes precautions on private entities as well, such as requiring banks to have sufficient capital reserves to withstand a financial crisis or mandating that airlines have emergency procedures in place.[44] A comprehensive review of these alternatives to

---

38. *See* Dillon & Tinsley, *supra* note 33, at 1436–38 (finding that near misses may decrease perceived risk).

39. For examples and factors influencing such calls, see, for example, Thomas A. Birkland & Megan K. Warnement, *Focusing Events, Risk & Regulation*, *in* POLICY SHOCK: RECALIBRATING RISK AND REGULATION AFTER OIL SPILLS, NUCLEAR ACCIDENTS, AND FINANCIAL CRISES 107, 118–20 (Edward J. Balleisen, Lori S. Bennear, Kimberly D. Krawiec & Jonathan B. Wiener eds., 2017) (noting that the greater the "perceived lack of control, dread, catastrophic potential, fatal consequences, and the inequitable distribution of risks and benefits . . . the more likely there will be greater attention to an event and call for regulatory action" (internal quotation marks omitted)).

40. Tversky & Kahneman, *supra* note 31, at 228.

41. *Id.* at 228–29.

42. The costs of global warming or a financial crisis are shared broadly. In contrast, the gains of oil drilling or bank investments concentrate in the executives of a few companies undertaking those activities. These market failures help explain the need for public imposition of risk identification on even economically rational private actors. *See, e.g.*, Richard Squire, *Shareholder Opportunism in a World of Risky Debt*, 123 HARV. L. REV. 1151, 1152 (2010) (showing how "seemingly reckless conduct . . . can in fact be fully rational from the perspective of shareholders").

43. *See, e.g.*, Robert F. Weber, *The Corporate Finance Case for Deliberation-Oriented Stress Testing Regulation,* 39 J. CORP. L. 833, 848–50 (2014) (identifying private-sector risk measurement tools such as value-at-risk, expected shortfall, risk-adjusted return on capital, scenario analysis, decision tree analysis, and Monte Carlo modeling).

44. *See* Fanto, *supra* note 17, at 736 (discussing the "growth in the importance of risk management" in "the regulation of capital"); Terry L. Turner, Note, *Carry-On Baggage—Are the Regulations Doing Their Job?*, 63 J. AIR L. & COM. 565, 569 (1998) (analyzing airline regulations

stress tests, and a comparison to each, is beyond the scope of this Article. However, a quick sketch of the administrative state's risk management trajectory will provide some helpful background for seeing stress tests' potential contributions to existing governance.

Agencies have long attempted to lessen risks by pervasively deploying inspectors, examiners, and other regulatory monitors to ensure that businesses are following the law.[45] In response to Civil War–era financial instability that threatened to undermine the North's military campaign, Congress ordered routine examinations of banks.[46] In 1907, federal inspections of meat manufacturers started after Upton Sinclair's *The Jungle* exposed vermin-infested facilities.[47]

Regulatory monitoring ultimately took a risk management turn, led by shifts in the private sector. In 1984, a poisonous gas leak at a plant in Bhopal, India, killed over three thousand people in nearby villages and injured many more.[48] The plant's owner, Union Carbide, was one of the world's largest businesses but closed down due to the incident.[49] Events such as the Bhopal gas leak spurred a new paradigm across diverse industries in which the law compelled businesses to develop risk management plans that regulators then audited and in many cases supplemented with their own risk assessments.[50]

For instance, authorized by the Clean Air Act,[51] the EPA promulgated rules requiring companies working with large quantities of toxic chemicals to submit a hazard analysis of ways that unintended

---

that "evolved under a discussion of problems associated with crashworthiness and emergency evacuation procedures").

45. *See* Rory Van Loo, *Regulatory Monitors: Policing Firms in the Compliance Era*, 119 COLUM. L. REV. 369, 384–86 (2019) (summarizing the history of regulatory monitoring).

46. National Bank Act of 1864, ch. 106, § 1, 13 Stat. 99 (codified as amended in scattered sections of 12 U.S.C.) (establishing the Office of the Comptroller of the Currency and providing it with authority to examine banks); ROSS M. ROBERTSON, THE COMPTROLLER AND BANK SUPERVISION: A HISTORICAL APPRAISAL 24–26 (1968) (providing a history of bank examinations).

47. *See* Meat Inspection Act, Pub. L. No. 59-242, 34 Stat. 1260 (1907) (codified at 21 U.S.C. §§ 601-695 (2012)); Roger Roots, *A Muckraker's Aftermath: The Jungle of Meat-Packing Regulation After a Century*, 27 WM. MITCHELL L. REV. 2413, 2417–19 (2001).

48. *See, e.g.*, Sukanya Pillay, *Absence of Justice: Lessons from the Bhopal Union Carbide Disaster for Latin America*, 14 MICH. ST. J. INT'L L. 479, 483–84 (2006).

49. *See id.*

50. *See, e.g.*, Coglianese & Lazer, *supra* note 15, at 725 ("[M]anagement-based regulation shifts the locus of policy decisionmaking from the government to private parties."); Bradley C. Karkkainen, *Information as Environmental Regulation: TRI and Performance Benchmarking, Precursor to a New Paradigm?*, 89 GEO. L.J. 257, 260 (2001) (depicting toxic releases inventory requirements as "a watershed, pioneering the systematic use of *performance monitoring* and *benchmarking* as regulatory tools" (footnote omitted)); Saule T. Omarova, *Wall Street as Community of Fate: Toward Financial Industry Self-Regulation*, 159 U. PA. L. REV. 411, 461 (2011) (summarizing moves in the chemical manufacturing, nuclear, and financial industries toward risk assessment and self-regulation).

51. Clean Air Act Amendments of 1990, Pub. L. No. 101-549, 104 Stat. 2399, 2571 (codified at 42 U.S.C. § 7412) (mandating that the EPA write risk management regulations).

release might occur. The companies must rank the possibilities by factors such as the number of people hypothetically contaminated.[52] The firms must then develop written operating procedures for emergency situations.[53] Agencies also model future risks of financial shocks, environmental catastrophes, or other threats.[54]

Yet regulatory monitoring checks whether firms are complying with existing legal requirements—meaning that those examinations de-bias best if the legal rule writers have sufficiently removed their own biases.[55] Moreover, observers have concluded that when both private and public actors create risk management plans, they tend to make familiar and comfortable assumptions,[56] in line with the psychological tendencies discussed above. The forecasting of an event makes the individual think about it, but (like having a near miss) that is not the same as experiencing the event.[57] By creating a false sense of security without addressing those psychological biases, regulatory monitoring and threat modeling can breed complacency.[58]

Stress tests aim to address the availability heuristic more directly than do regulatory monitoring and risk forecasting. Role-playing simulations would do this by causing people to experience the consequences of a rare emergency that may not otherwise be top of mind, thereby making it more likely that participants will later recall that threat when planning. Like an accelerating treadmill exploring the outer bounds of a patient's heart health, an administrative stress test aims to push an organization beyond its comfort zone to uncover complacency and prompt change.

B. Case Studies in Current Stress Tests

This Section considers three areas in which agencies run stress tests: public health, the environment, and finance. These three areas

52. 40 C.F.R. § 68 (2007).

53. *Id.*

54. *See, e.g.*, Baradaran, *supra* note 9, at 1250; Fanto, *supra* note 17, at 733–35 (analyzing environmental law risk management); Judge, *supra* note 17, at 659–61.

55. *See* Van Loo, *supra* note 45, at 399 (discussing risk management plans in the context of regulatory monitoring).

56. *See, e.g.*, Stephen J. Choi & A.C. Pritchard, *Behavioral Economics and the SEC*, 56 STAN. L. REV. 1, 72 (2003) ("Heuristics play a large role in how regulators make decisions."); Cohen, *supra* note 17, at 390–91 (observing shortcomings in risk analyses).

57. *See supra* Section I.A.1.

58. *See, e.g.*, Cohen, *supra* note 17, at 390–91 ("Reliance on risk assessment and risk management discourses also can induce unwarranted complacency and encourage excessive risk-taking."); *see also* Jerome H. Powell, Chair, Bd. of Governors of the Fed. Rsrv. Sys., Welcoming Remarks at Stress Testing: A Discussion and Review (Jul. 9, 2019) (observing that routine examinations can breed complacency).

are a subset of public stress tests, which occur at all levels of federal, state, and local government in areas ranging from power grid failures to fire prevention.[59] Although stress tests in different areas may have distinct dynamics, these case studies illustrate both what stress tests offer and how they are limited as currently implemented.

1. Public Health

For years, the Department of Health and Human Services ("HHS") has attempted to focus health care actors' minds on catastrophes by mandating an annual "coalition surge test."[60] HHS requires hospital groups and public health agencies to participate in these tests as a condition of receiving federal funding.[61] Prompted by events such as the 9/11 terrorist attacks and Hurricane Katrina, which exceeded hospitals' capacities, these tests assess local funding recipients' "ability to work in a coordinated way using their own systems and plans to find appropriate destinations for patients" during an unprecedented need for evacuations.[62]

In 2019, HHS also voluntarily completed an ambitious pandemic stress test. The Spanish Flu of 1918 infected about one-third of the global population and killed more people than World War I.[63] For years before COVID-19 arrived, "[v]irtually all experts agree[d] that it is not a question of *if* but *when* another influenza pandemic as deadly as the

---

59. *See, e.g.*, *Preparing for a Catastrophe: The Hurricane Pam Exercise: Hearing Before the S. Comm. on Homeland Sec. and Governmental Affs.*, 109th Cong. 1, 3 (2006) [hereinafter *FEMA Simulation Comm. Hearing*] (statement of Susan M. Collins, Chairman, S. Comm. on Homeland Sec. & Governmental Affs.) (mentioning the breadth of simulations); Neal H. Lewis, *Interpreting the Oracle: Licensing Modifications, Economics, Safety, Politics, and the Future of Nuclear Power in the United States*, 16 ALB. L.J. SCI. & TECH. 27, 46 (2006) (explaining how a simulation indicated that an Oak Ridge National Laboratory plant should be shut down because a power blackout would generate lethal, uncontainable radiation); *see also* PAC. GAS & ELEC. CO., AMENDED 2019 WILDFIRE SAFETY PLAN 34 (Feb. 2019), https://www.pge.com/pge_global/common/pdfs/safety/emergency-preparedness/natural-disaster/wildfires/Wildfire-Safety-Plan.pdf [https://perma.cc/XM9T2G87] (summarizing an exercise in which local traffic authorities and the responsible utility company launched simulations in high-risk areas to ensure adequate evacuation routes and procedures).

60. OFF. OF THE ASSISTANT SEC'Y FOR PREPAREDNESS & RESPONSE, U.S. DEP'T OF HEALTH & HUM. SERVS., 2017-2022 HOSPITAL PREPAREDNESS PROGRAM: PERFORMANCE MEASURES IMPLEMENTATION GUIDE 5 (2017), https://www.phe.gov/Preparedness/planning/hpp/reports/Documents/hpp-pmi-guidance-2017.pdf [https://perma.cc/5ZR7-B9X6] [hereinafter HHS PERFORMANCE MEASURES].

61. *Id.*

62. *Id.* at 32.

63. Andrew Burns, Dominique van der Mensbrugghe & Hans Timmer, Evaluating the Economic Consequences of Avian Influenza (2008) (unpublished manuscript), https://www.researchgate.net/publication/237345628_Evaluating_the_Economic_Consequences_of_Avian_Influenza1 [https://perma.cc/JFZ2-VQTG] (originally published in slightly different form in WORLD BANK, GLOBAL DEVELOPMENT FINANCE: THE DEVELOPMENT POTENTIAL OF SURGING CAPITAL FLOWS 36 (2006)).

Spanish Flu will occur."[64] Forecasts predicted that if such a pandemic materialized, it could result in at least two hundred million deaths, including millions in the United States alone.[65] The expected economic costs were also substantial, with the Congressional Budget Office estimating that a modern pandemic like the Spanish Flu would cost the world over $3 trillion.[66]

In light of those predictions, and following close calls with Ebola and other outbreaks, the Obama Administration began planning ambitious role-plays with federal, state, and local officials testing the nation's readiness for pandemics.[67] Those efforts culminated in a 2019 exercise under the Trump Administration, codenamed "Crimson Contagion."[68] Managed by HHS, officials from the Centers for Disease Control and Prevention ("CDC") and other agencies role-played—through conference calls and meetings—their coordinated response to a pandemic that infected 110 million Americans.[69]

The simulation lasted eight months in four phases, and involved participants from twelve federal agencies, twelve states, ninety-six local jurisdictions, eighty-seven hospitals, and over one hundred private-sector partners.[70] The script released information to participants in real time, informing them early on that over twelve thousand cases had been reported in the United States.[71] Role-play participants would then respond, such as by calling for social distancing and working from home after being informed of a first wave of U.S. cases.[72]

---

64. Vickie J. Williams, *Fluconomics: Preserving Our Hospital Infrastructure During and After a Pandemic*, 7 YALE J. HEALTH POL'Y L. & ETHICS 99, 100 (2007); *see also* Christina Y. Chan, *Support for the First Line of Defense in Public Health Emergencies*, 105 NW. U. L. REV. 1347, 1348–49 (2011) ("Most public health officials agree that it is simply a matter of time before the United States suffers a disastrous influenza epidemic.").

65. Burns et al., *supra* note 63.

66. *Id.*; *see also* ROBERT ARNOLD, JEANNE DE SA, TIM GRONNIGER, ALLISON PERCY & JULIE SOMERS, CONG. BUDGET OFFICE, A POTENTIAL INFLUENZA PANDEMIC: POSSIBLE MACROECONOMIC EFFECTS AND POLICY ISSUES 44 (July 2006), https://www.cbo.gov/sites/default/files/109th-congress-2005-2006/reports/12-08-birdflu.pdf [https://perma.cc/4Z6H-PJP8].

67. David E. Sanger, Eric Lipton, Eileen Sullivan & Michael Crowley, *Before Virus Outbreak, a Cascade of Warnings Went Unheeded*, N.Y. TIMES, https://www.nytimes.com/2020/03/19/us/politics/trump-coronavirus-outbreak.html (last updated Sept. 4, 2021) [https://perma.cc/SK5C-MEF3].

68. *See* OFF. OF THE ASSISTANT SEC'Y FOR PREPAREDNESS & RESPONSE, U.S. DEP'T OF HEALTH & HUM. SERVS., CRIMSON CONTAGION 2019 FUNCTIONAL EXERCISE: DRAFT AFTER-ACTION REPORT 5, app. A (Oct. 2019), https://int.nyt.com/data/documenthelper/6824-2019-10-key-findings-and-after/05bd797500ea55be0724/optimized/full.pdf#page=1 [https://perma.cc/8SY9-HX4R] [hereinafter HHS PANDEMIC SIMULATION DRAFT REP.] (unpublished draft).

69. *Id.*

70. *Id.* at 2–5, app. A.

71. Sanger et al., *supra* note 67.

72. *See* HHS PANDEMIC SIMULATION DRAFT REP., *supra* note 68, at 23.

HHS produced a report at the end of the simulation that suggested the U.S. public health infrastructure was unprepared for a major pandemic. The lack of preparedness identified fell into two main infrastructural categories of note: hospitals and laboratories.

In terms of health care, the Crimson Contagion report identified the challenge of shortages for ventilators, needles, syringes, and other supplies.[73] The United States later faced these same problems in responding to the COVID-19 pandemic. For instance, adequate health care infrastructure for a pandemic requires surge capacity—the ability to accommodate a huge increase in persons requiring care.[74] Yet doctors, nurses, and other crucial medical personnel lacked adequate masks during early COVID-19 surges. This shortage caused more health care workers to become sick and die and thus deprived the health care system of sufficient support in some areas.[75]

Access to a hospital bed with a functioning ventilator has made the difference between life and death for many patients struck with COVID-19-induced pneumonia.[76] Yet the nation had only a fraction of the ventilators and beds needed in some areas, leaving many without lifesaving treatment.[77]

Finally, scientific infrastructure means, among other things, virus screening (testing). A key failure of Crimson Contagion was not considering the nation's capacity for ramping up screening.[78] Screening patients is essential for slowing the spread of a pandemic because it allows officials to identify, trace, and quarantine those who are

---

73. *Id.* at 39.

74. Williams, *supra* note 64, at 132.

75. Kent Babb, Brittany Shammas & Ariana Eunjung Cha, *Hundreds of Health-Care Workers Lost Their Lives Battling the Coronavirus, Now a Series of 'What Ifs' Haunt Their Friends and Colleagues.*, WASH. POST (June 17, 2020), https://www.washingtonpost.com/graphics/2020/health/healthcare-workers-death-coronavirus/?itid=lk_inline_manual_30 [https://perma.cc/R7VF-9DRD] ("[For healthcare workers, s]imply reporting for work can mean gambling with your life, and the odds grow longer when masks and other personal protective equipment become difficult to get.").

76. *Interim Clinical Guidance for Management of Patients with Confirmed Coronavirus Disease (COVID-19)*, CDC, https://www.cdc.gov/coronavirus/2019-ncov/hcp/clinical-guidance-management-patients.html (last updated Feb. 16, 2021) [https://perma.cc/L8Z5-AGJE] [hereinafter *COVID-19 Interim Clinical Guidance*].

77. *See, e.g.*, OFF. OF INSPECTOR GEN., U.S. DEP'T OF HEALTH & HUM. SERVS., OEI-06-20-00300, HOSPITAL EXPERIENCES RESPONDING TO THE COVID-19 PANDEMIC: RESULTS OF A NATIONAL PULSE SURVEY MARCH 23-27, 2020 (2020), https://oig.hhs.gov/oei/reports/oei-06-20-00300.pdf [https://perma.cc/LVU3-FEQ6] (describing how testing shortages created a lack of bed space and lack of personal protective equipment); Nicholas Kulish, Sarah Kliff & Jessica Silver-Greenberg, *The U.S. Tried to Build a New Fleet of Ventilators. The Mission Failed.*, N.Y. TIMES, https://www.nytimes.com/2020/03/29/business/coronavirus-us-ventilator-shortage.html (last updated Apr. 20, 2020) [https://perma.cc/HV7B-CEPS] (discussing massive shortage of ventilators in face of public health crises).

78. HHS PANDEMIC SIMULATION DRAFT REP., *supra* note 68, at 46.

infected—the most effective strategy for significantly reducing mortality.[79]

Despite the importance of screening, the United States suffered from a critical screening capacity shortage for COVID-19 in 2020, leaving even many high-risk patients untested.[80] A rigid and slow approval process was part of the problem for lab facilities that sought to screen.[81] For months, tens of thousands of research and health facilities across the United States remained dormant even though they had the capacity to screen in large numbers.[82]

Some caution is warranted for quality control and avoiding false positives. However, better prepared countries successfully balanced these concerns with dramatically faster responses. The Netherlands and South Korea, for instance, conducted more than three hundred times the number of tests per capita than the United States in the first weeks of the outbreak.[83] Although many factors go into morbidity, these countries also had forty percent and ninety-eight percent lower deaths per capita than the United States, respectively.[84] As one epidemiology professor described it, "The lack of testing in the United States is a debacle . . . . We're supposed to be the best biomedical powerhouse in

79. NEIL M. FERGUSON ET AL., IMPERIAL COLL. COVID-19 RESPONSE TEAM, IMPACT OF NON-PHARMACEUTICAL INTERVENTIONS (NPIS) TO REDUCE COVID-19 MORTALITY AND HEALTHCARE DEMAND 1 (Mar. 2020), https://www.politico.eu/wp-content/uploads/2020/03/Imperial-College-COVID19-NPI-modelling-16-03-2020.pdf [https://perma.cc/8BTR-KVQ6] (identifying quarantine as the single most important way to reduce mortality); Brian J. Coburn, Bradley G. Wagner & Sally Blower, *Modeling Influenza Epidemics and Pandemics: Insights into the Future of Swine Flu (H1N1)*, BMC MED., June 22, 2009, at 1, 4 (discussing the effectiveness of "household quarantine" in reducing spread of outbreak).

80. *See* Andrew Ryan, Kay Lazar & Liz Kowalczyk, *Severe Shortage of Tests Blunts Coronavirus Response, Boston Doctors Say*, BOS. GLOBE, https://www.bostonglobe.com/2020/03/12/nation/severe-shortage-tests-blunts-coronavirus-response-boston-doctors-say (last updated Mar. 12, 2020, 12:51 PM) [https://perma.cc/K6LV-PRQY].

81. *Id.*; *see also infra* Section III.A. Another challenge is having a sufficient amount of the chemical needed for the test—a chemical that is not specific to coronavirus and thus could have been held in reserve. *See, e.g.*, Ryan et al., *supra* note 80 (describing testing shortfalls).

82. *See, e.g.*, *United States 2017 Laboratory Market Report - Research and Markets*, BUS. WIRE (Sept. 29, 2017, 5:31 AM), https://www.businesswire.com/news/home/20170929005241/en/United-States-2017-Laboratory-Market-Report [https://perma.cc/M4F5-4T99] (reporting 121,973 laboratory facilities in United States); *see also* Ryan et al., *supra* note 80 (explaining the slow approval testing mobilization).

83. Aylin Woodward & Skye Gould, *One Chart Shows How Many Coronavirus Tests Per Capita Have Been Completed in 8 Countries. The US Is Woefully Behind.*, BUS. INSIDER (Mar. 9, 2020, 7:20 PM), https://www.businessinsider.com/coronavirus-testing-covid-19-tests-per-capita-chart-us-behind-2020-3 [perma.cc/U22U-GZCE] (putting the ratio of tests per capita between other countries and the United States at about 700 for South Korea, 347 for the U.K., 350 for the Netherlands, and 401 for Israel).

84. *See Coronavirus (COVID-19) Deaths Worldwide Per One Million Population as of September 28, 2021, by Country*, STATISTA (Sept. 5, 2021), https://www.statista.com/statistics/1104709 [perma.cc/5U4V-854N].

the world and we've been unable to do something that every other country has been able to do."[85]

Under withering public criticism, the FDA eventually developed an accelerated approval process for labs seeking to run diagnostics for COVID-19.[86] In the meantime, however, without the advanced warning systems that the Crimson Contagion simulation could have identified as necessary,[87] a deadly virus silently spread for at least six weeks in communities nationwide.[88]

Despite the impossibility of knowing how many fatalities could have been avoided, the question remains of why even some low-cost organizational steps were not taken before the pandemic, such as developing a rapid screening approval process that could be immediately implemented in the case of emergency. The discussion below will return to the issue of why Crimson Contagion failed to prompt more action—why it failed to overcome the psychological barriers to effective crisis management. For now, the main point is that despite the simulation and prominent predictions of inevitable large-scale contagion, the United States public health system was underprepared in ways that were or could have been identified by a stress test in advance.

2. The Environment

Environmental and emergency response agencies commonly role-play hazards. Those simulations may respond to environmental catastrophes caused by either nature or industry. For natural disasters, at various times throughout U.S. history, a single wildfire, drought, heat wave, or earthquake has taken thousands of American lives and cost tens of billions of dollars in damage.[89] Fires are growing in intensity, but in recent decades storms have caused the greatest

85. Ryan et al., *supra* note 80 (quoting Harvard Professor of Epidemiology Marc Lipsitch).

86. Press Release, U.S. Food & Drug Admin., Coronavirus (COVID-19) Update: FDA Issues New Policy to Help Expedite Availability of Diagnostics (Feb. 29, 2020), https://www.fda.gov/news-events/press-announcements/coronavirus-covid-19-update-fda-issues-new-policy-help-expedite-availability-diagnostics [https://perma.cc/8L3M-9BPE].

87. *Infra* Section III.A.

88. *Infra* Section III.A (mentioning experts suspecting that further research will likely reveal that the virus was spreading silently even earlier than six weeks); *see also* OFF. OF INSPECTOR GEN., *supra* note 77, at 5 (discussing how a continued lack of quick testing caused hospital overcrowding due to inability to move patients out of isolated wards until they were cleared of having the virus).

89. *Billion-Dollar Weather and Climate Disasters: Events*, NAT'L OCEANIC & ATMOSPHERIC ADMIN., https://www.ncdc.noaa.gov/billions/events (last visited Feb. 17, 2022) [https://perma.cc/5ZVV-YPL4].

destruction.[90] After storms strike, FEMA coordinates national, regional, state, and local relief.[91] Despite the frequency and magnitude of extreme storms, and hurricanes in particular, the agency has a mixed record with them.[92]

In the late 1990s the state of Louisiana requested a simulation of a catastrophic hurricane.[93] FEMA approved the request because after several close calls of major hurricanes hitting New Orleans, it knew that the city was susceptible to devastating flooding.[94] However, the agency lacked a sense of urgency, as would be expected from the psychological research into how people process close calls.[95] FEMA repeatedly pushed back the planned start date, ultimately delaying it over five years, and during that time cut several proposed scenarios.[96] When it ultimately happened in 2004, the exercise was still substantial, bringing together about three hundred emergency officials from all levels of state and federal government to decide how they would handle swamped hospitals, many deaths, and hundreds of thousands of residents trapped in flooded areas.[97]

The following year, Hurricane Katrina struck New Orleans. Despite a recently completed simulation that had assumed thousands of deaths, FEMA delayed sending rescue workers and supplies.[98] As a result, many of the over one million displaced local residents went without food and aid, with some stranded on the roofs of their submerged homes.

The agency also botched the coordination of nonprofits, businesses, and government entities by withholding key permissions needed to mobilize many who stood ready to help with food and medical supplies.[99] For instance, FEMA turned away truckloads of bottled water

---

90. *Id.* Wildfires are second in the past few years and, although unlikely, if that trajectory continues could eventually rival hurricanes. *Id.*

91. *See, e.g.*, Yishai Blank & Issi Rosen-Zvi, *Reviving Federal Regions*, 70 STAN. L. REV. 1895, 1942 (2018) (discussing FEMA's organizational structure).

92. *See* Yxta Maya Murray, *"FEMA Has Been A Nightmare:" Epistemic Injustice in Puerto Rico*, 55 WILLAMETTE L. REV. 321, 325 (2019) (discussing the tolls of Hurricanes Maria and Katrina).

93. SELECT BIPARTISAN COMM. TO INVESTIGATE THE PREPARATION FOR AND RESPONSE TO HURRICANE KATRINA, A FAILURE OF INITIATIVE, H.R. REP. NO. 109-377, at 2, 13 (2d Sess. 2006) (emphasizing testimony that requested funding to implement necessary steps was denied and admitting that levees were not built to withstand the most severe hurricanes); *see also infra* Section II.B.

94. H.R. REP. NO. 109-377, *supra* note 93, at 2, 13.

95. *See supra* Section I.A (discussing the organizational psychology of crisis management).

96. *FEMA Simulation Comm. Hearing*, *supra* note 59, at 1, 3 (statement of Susan M. Collins, Chairman, S. Comm. on Homeland Sec. & Governmental Affs.).

97. *Id.* at 2, 17.

98. *Id.*

99. *Id.* at 17.

donated by Walmart even as many residents struggled to obtain clean water and did not allow the Coast Guard to deliver one thousand gallons of fuel, which was in short supply.[100] After arriving, firefighters spent days waiting in their hotels rather than rescuing people.[101]

These deficiencies were rooted in the absence of a detailed FEMA hurricane response plan in advance, which meant that the agency needed to devote precious time during the crisis to developing operating procedures.[102] The Governor of Louisiana summed up his experience in dealing with FEMA at the time by noting, "We wanted soldiers, helicopters, food and water . . . . They wanted to negotiate an organizational chart."[103]

Hurricane Katrina lives in U.S. history as one of the costliest and deadliest natural disasters, from which New Orleans still has not economically recovered fifteen years later.[104] The FEMA simulation had uncovered many of these problems, but missed others because of the agency's decision to reduce the scenarios considered.[105] Regardless, the delays in completing the exercise meant that the simulation's final report did not arrive sufficiently in advance to make meaningful organizational changes that might have lessened the decimation.[106]

3. The Financial System

After the economy teetered on the edge of a cliff in 2008, with the Dow Jones Industrial Average down by over fifty percent, U.S. Treasury Secretary Timothy Geithner announced that the first step to prevent another crisis would be requiring "banking institutions to go through a carefully designed comprehensive stress test, to use the medical term."[107] The Dodd-Frank Act of 2010 mandated that a "systemically important financial institution"—an entity whose

100. Scott Shane, Eric Lipton & Christopher Drew, *After Failures, Government Officials Play Blame Game*, N.Y. TIMES (Sept. 5, 2005), https://www.nytimes.com/2005/09/05/us/nationalspecial/after-failures-government-officials-play-blame-game.html [https://perma.cc/P2UM-Y57H].

101. Greg Bluestein, *Hundreds of Firefighters Stuck at Georgia Hotel for Days Awaiting Orders*, Lincoln J. Star (Sept. 7, 2005), https://journalstar.com/special-section/news/hundreds-of-firefighters-stuck-at-georgia-hotel-for-days-awaiting-orders/article_8db0b5a5-d7ff-501b-9f69-5168c4cb55ff.html [https://perma.cc/K9UW-PLKT].

102. *See generally* David A. Super, *Against Flexibility*, 96 CORNELL L. REV. 1375, 1451–53 (2011) (explaining FEMA's planning delays).

103. *See* Shane et al.*, supra* note 100 (internal quotation marks omitted).

104. *Id.*

105. *See infra* Section I.C.

106. *FEMA Simulation Comm. Hearing*, *supra* note 59, at 2.

107. *See* Press Release, U.S. Dep't of the Treasury, Secretary Geithner Introduces Financial Stability Plan (Feb. 10, 2009), http://www.treasury.gov/press-center/press-releases/Pages/tg18.aspx [https://perma.cc/H9LE-QSHC].

collapse would endanger the economy, like a big bank—undertake annual stress tests supervised by a regulator.[108] These tests create fictitious simulations to help regulators understand how financial institutions "will respond to severe, yet plausible, stressed market conditions such as low economic output, high unemployment, [and] stock market crashes."[109] The goal is not to punish the bank for failing.[110] Instead, the regulator might require a bank that failed the test to take preventive steps, such as lowering its risky debt.[111]

The Federal Reserve is the primary supervisor of those tests and conducts some stress tests of banks on its own.[112] The focal point is a computer simulation of how a bank's balance sheet would respond to various circumstances, such as a deep recession that drives unemployment up to thirteen percent and cuts the stock market's value in half.[113] Most importantly, they determine whether the bank would have sufficient funds to continue operations. Banks have consistently passed these exercises in recent years, giving observers greater confidence in the financial system's resilience during the COVID-19 downturn.[114] Stress tests are seen as an "indispensable tool of any systematic effort to reduce the likelihood and impact of crisis events."[115]

Precautionary measures are crucial because a single large bank failure can spark a chain reaction that brings down the entire economy.[116] Just as doctors probe hearts to prevent medical

---

108. Dodd-Frank Wall Street Reform and Consumer Protection Act of 2010 § 165(i), Pub. L. No. 111-203, 124 Stat. 1430 (codified at 12 U.S.C. § 5365) (requiring stress tests for bank holding companies with over $10 billion in assets). For decades, stress tests existed in finance, mostly under the internal control of banks. *See, e.g.*, MICHEL CROUHY, DAN GALAI & ROBERT MARK, RISK MANAGEMENT 232–40 (2001) (describing the use of stress tests).

109. Weber, *supra* note 43, at 833 (mentioning also liquidity shortages, high default rates, and failures of large counterparties).

110. *Id.*

111. MICHAEL S. BARR, HOWELL E. JACKSON & MARGARET E. TAHYAR, FINANCIAL REGULATION: LAW AND POLICY 189, 265 (2016).

112. *Id.*

113. Daniel K. Tarullo, Governor, Bd. of Governors of the Fed. Rsrv. Sys., Speech at the Federal Reserve Bank of Chicago Annual Risk Conference: Developing Tools for Dynamic Capital Supervision (Apr. 10, 2012), https://www.federalreserve.gov/newsevents/speech/tarullo20120410a.htm [https://perma.cc/8UHC-M7PH].

114. *See, e.g.*, BD. OF GOVERNORS OF THE FED. RSRV. SYS., DODD-FRANK ACT STRESS TEST 2019: SUPERVISORY STRESS TEST RESULTS (June 2019), https://www.federalreserve.gov/publications/files/2019-dfast-results-20190621.pdf [https://perma.cc/S3BR-LD33]; Neil Irwin, *Something Weird Is Happening on Wall Street, and Not Just the Stock Sell-Off*, N.Y. TIMES (Mar. 12, 2020), https://www.nytimes.com/2020/03/12/upshot/markets-weird-coronavirus.html [https://perma.cc/A5YJ-QEMR] (expressing greater confidence).

115. Giovanni Carosio, *Foreword* to STRESS-TESTING THE BANKING SYSTEM, at xxi (Mario Quagliariello ed., 2009).

116. *See* Steven L. Schwarcz, *Systemic Risk*, 97 GEO. L.J. 193, 193 (2008).

emergencies and deaths, so must regulators probe vital financial institutions to prevent the economy from collapsing.

Stress tests are necessary because triggers for financial instability have repeatedly caught regulators unaware in recent decades. The most recent example is the 2007 mortgage crisis, in which U.S. financial institutions aggressively extended loans to home buyers, too often deceiving low-income applicants.[117] Many borrowers and lenders were also overly optimistic because they assumed housing prices would continue to rise.[118] By 2008, housing prices began to fall and borrowers became increasingly unable to pay their mortgages, resulting in mass defaults and over five million people losing their homes.[119] Some banks had extended many mortgages that would not be repaid and the banks were also exposed to highly leveraged mortgage-backed securities—an investment device that gambled on people's ability to pay back their mortgages.[120]

Regulators missed this threat despite the fact that one of their core jobs is to ensure that banks do not have too much risky debt, and to that end they routinely examine banks' records.[121] Indeed, a team of regulators called "resident inspectors" is on-site year-round at the largest banks combing through their activities to ensure compliance with the law.[122] Although regulators knew that banks depended on mortgages as safe and reliable debt instruments before 2008, examiners failed to realize that much of that debt was at risk of not being repaid.[123] While the full causes of the subsequent financial collapse and recession are still debated, the mortgage crisis contributed substantially.[124]

As another example, when the 2006 avian flu surge in Asia stoked worldwide alarm, an interagency group of U.S. financial regulators instructed banks to include pandemics in their operational plans.[125] However, because the outbreak was ultimately milder than originally thought, the psychology of near misses made it unlikely that regulators would increase their concern about

---

117. *See* Steven L. Schwarcz, *Markets, Systemic Risk, and the Subprime Mortgage Crisis*, 61 SMU L. REV. 209, 210 (2008).

118. *See* Patricia A. McCoy & Susan M. Wachter, *Why the Ability-to-Repay Rule is Vital to Financial Stability*, 108 GEO. L.J. 649, 652 (2020).

119. *Id.*

120. *Id.*

121. *See, e.g.*, Van Loo, *supra* note 45, at 404, 427 (summarizing financial regulatory examination of banks).

122. *Id.* at 412.

123. *See id.* at 371–72.

124. *See id.* at 395.

125. *See* FED. FIN. INSTS. EXAMINATION COUNCIL, INTERAGENCY STATEMENT ON PANDEMIC PLANNING 5 (2020), https://www.ffiec.gov/press/pandemicguidance.pdf [https://perma.cc/8SSZ-X489].

pandemics. Moreover, as predicted by the availability heuristic, pandemics had faded into the background of agency concerns by the time COVID-19 arrived fourteen years later.[126]

As a result, financial regulators were slow to act during the COVID-19 crisis. By late January 2020, the World Health Organization had declared an international public health emergency,[127] President Trump had begun to restrict travel from China,[128] and the Secretary of Commerce wrote a memo warning that the virus could cost the United States trillions of dollars and endanger millions of Americans' lives.[129] Yet financial regulators waited until March 6, a week after a stock market wiped out over three trillion dollars in value, to push banks to "update" their pandemic operational plans.[130] That update was the first in over a decade, and called for considering the risks of how to continue operations remotely or with forty percent bank-worker absenteeism, and how to reassess instability risks.[131]

Financial regulators' long inattention to pandemics means that the financial system was not tested for the kind of economic shutdown caused by COVID-19, and as a result regulators were largely in the dark about how financial institutions would fare.[132] Fortunately, since the resulting recession wound up being short-lived, the lack of financial regulation preparation did not have severe consequences.

Financial regulators have a particularly difficult prediction task. Whereas pandemics will originate within a public health regulator's field of expertise, the catalyst for a financial crisis may lie outside finance—coming from housing, health, or political unrest. Nonetheless, it is concerning that the potential triggers of financial upheaval have consistently surprised regulators even when regulators directly oversaw the products and institutions (such as mortgages and banks) that ultimately caused the collapse.[133] Despite the most extensive

126. *See* Section I.A (discussing the psychology of near misses and availability).

127. Press Briefing, World Health Org. [WHO], Coronavirus Emergency Committee Second Meeting (Jan. 30, 2020), https://www.who.int/docs/default-source/coronaviruse/transcripts/ihr-emergency-committee-for-pneumonia-due-to-the-novel-coronavirus-2019-ncov-press-briefing-transcript-30012020.pdf [https://perma.cc/8DK7-MSY6] (providing transcript).

128. Proclamation No. 9984, 85 Fed. Reg. 6709 (Jan. 31, 2020).

129. Maggie Haberman, *Trade Adviser Warned White House in January of Risks of a Pandemic*, N.Y. TIMES, https://www.nytimes.com/2020/04/06/us/politics/navarro-warning-trump-coronavirus.html (last updated Mar. 31, 2021) [https://perma.cc/D4WZ-P3KY].

130. FED. FIN. INSTS. EXAMINATION COUNCIL, *supra* note 125, at 1.

131. *Id.* at 6.

132. *See infra* Section II.C (discussing limits of financial regulators' planning).

133. Consider, for example, how law professors Kathleen Engel and Patricia McCoy spoke of the mortgage crisis several years before the financial crisis—at a time when it was still possible to prevent much of the devastation that resulted. *See* Kathleen C. Engel & Patricia A. McCoy, *A Tale of Three Markets: The Law and Economics of Predatory Lending*, 80 TEX. L. REV. 1255, 1257–59

adoption of external stress tests in these three case studies, financial regulators have yet to show that they are ready for "the complexities of modern financial markets and investment securities" that present "perhaps the greatest financial-market challenge of the future."[134]

C. Existing Stress Test Limitations

Even with their identifiable flaws and subsequent administrative shortcomings, it is still possible that stress tests are helping to prevent even worse results. Additionally, there is a risk of hindsight bias in assessing the response to stress tests. Given limited resources, agency leaders cannot prepare for all conceivable threats. Nonetheless, the above case studies indicate that as currently implemented stress tests have several potentially significant limitations worth examining further: voluntariness, confidentiality, a lack of imagination, and the absence of comprehensive evaluation.

First, stress tests of administrative agencies are voluntary. Lawmakers have only imposed stress tests on private-sector entities, most notably banks. The Federal Reserve is under no obligation to simulate the regulatory decisions it would make in the face of big bank failures, a depression, or other financial shocks.[135]

Although HHS compels many state and local health care entities to complete a surge capacity hospital readiness exercise, no law demands something similar of HHS or FEMA.[136] HHS voluntarily initiated its Crimson Contagion exercise without any congressional oversight or mandate.[137] Similarly, although the CDC participated in Crimson Contagion, it is not obligated to undergo any pandemic simulation of its own.

(2002) (discussing the predatory tactics of subprime mortgage lenders and proposing that a "duty of suitability" be imposed on such lenders to incentivize efficient markets).

134. Steven L. Schwarcz, *Regulating Complexity in Financial Markets*, 87 WASH. U. L. REV. 211, 211 (2009).

135. Instead, the Federal Reserve views the stress tests of banks as a test of its own performance. *See* Randal K. Quarles, Vice Chair for Supervision, Bd. of Governors of the Fed. Rsrv. Sys., Remarks at a Research Conference Sponsored by the Federal Reserve Bank of Boston: Stress Testing: A Discussion and Review (July 9, 2019), https://www.federalreserve.gov/newsevents/speech/files/quarles20190709a.pdf [https://perma.cc/M7UY-MPWT].

136. *See* HHS PERFORMANCE MEASURES, *supra* note 60, at 32–33 (explaining the voluntary nature of the tests). A recent proposal would expand stress testing of hospitals. *See* Barak D. Richman & Steven L. Schwarcz, *Macromedial Regulation*, 82 OHIO ST. L.J. 727, 743 (2021).

137. *See Flu Season: U.S. Public Health Preparedness and Response: Written Testimony Before the Subcomm. on Oversight & Investigations of the H. Comm. on Energy & Com.*, 116th Cong. 8–9 (2019) (statement of Robert Kadlec, Assistant Secretary for Preparedness and Response, United States Department of Health and Human Services) (describing the breadth of the Crimson Contagion 2019 Functional Exercise).

The voluntary nature of the exercises can have consequences. For example, FEMA was free to follow any timetable it wanted in conducting its Louisiana hurricane simulation, unlike the mandatory annual stress tests of banks. The five-year delay proved costly, as necessary planning documents were not "generated before they were so urgently needed" for Katrina.[138] Moreover, the voluntary nature means that agency heads typically opt out of participating, leaving the exercises instead for midlevel or frontline employees.[139] This makes it less likely that organizational change will result.

Second, the public and sometimes Congress lack visibility into stress tests. Regulators usually do not publicly release the results. HHS kept the preliminary findings of Crimson Contagion nonpublic until an internal draft document, marked "DO NOT DISTRIBUTE," was leaked to the press during the COVID-19 pandemic.[140] The results of FEMA's hurricane exercise also remained private.[141] This secrecy can undermine accountability and makes it less likely public concern will pressure an agency to learn from the test's findings.

Third, as one would expect based on the psychological literature,[142] there is sometimes an imagination deficit in predicting the nature or scale of the next crisis. For example, Crimson Contagion did not consider whether agencies were positioned to ramp up nationwide virus testing, thus missing an actionable issue that could have dramatically mitigated the COVID-19 pandemic early on.[143]

Even in financial regulation, where stress tests have a longer track record and were designed to be forward looking, agencies have mostly used them to ensure banks can withstand a recurrence of prior shocks. For instance, stress tests now consider banks' mortgage exposure and ability to withstand something like the recession of 2008.[144] However, financial regulators never implemented pandemic economic shocks into their tests, despite a parade of near misses in Severe Acute Respiratory Syndrome (2002), bird flu (2006), swine flu (2009), and Middle East Respiratory Syndrome (2012).[145]

---

138. *FEMA Simulation Comm. Hearing, supra* note 59, at 2.

139. *See, e.g.*, *id.* at 20 ("[T]he effectiveness of the [hurricane simulation] was greatly diminished by the poor attendance of key decisionmakers . . . .").

140. Sanger et al., *supra* note 67.

141. *See FEMA Simulation Comm. Hearing*, *supra* note 59, at 1–2.

142. *See supra* Section I.A.

143. *See* HHS PANDEMIC SIMULATION DRAFT REP., *supra* note 68 (failing to mention CDC testing).

144. *See* Tarullo, Speech at the Federal Reserve Bank of Chicago Annual Risk Conference, *supra* note 113.

145. *See* Margo A. Bagley, *The Morality of Compulsory Licensing as an Access to Medicines Tool*, 102 MINN. L. REV. 2463, 2492 n.122 (2018) (listing the string of health challenges); Min Joo

Indeed, when the Federal Reserve released its stress test hypotheticals in February of 2020, even after the globe was on notice that COVID-19 had begun to ravage Wuhan, China, at an alarming rate, the "severely adverse scenario" did not reference a pandemic.[146] Moreover, some of those announced parameters were overly optimistic. The Federal Reserve assumed in its worst-case scenario that unemployment would rise to six percent, whereas during the COVID-19 crisis it surpassed fourteen percent; and its stress test assumed a drop in GDP of about ten percent whereas the actual drop was greater than thirty percent.[147]

Financial regulators' backward-looking stress tests target the availability heuristic for emergencies that tend to repeat themselves in similar forms—which is extremely valuable and may have prevented many financial crashes that would have otherwise occurred in similar forms.[148] However, this approach is less relevant to disruptions that may come in new forms or vary greatly in severity. That is particularly concerning because observers have long recognized that the "trigger to the next crisis will not be the same as the trigger to the last one—but there will be another crisis."[149]

A final limitation is that even when simulations are predictive, they often fail to evaluate the organization. To encourage agency leaders to effect organizational change, or for Congress to allocate more resources, stress tests would need to produce realistic assessments of unpreparedness. Yet candid evaluations are in tension with agency leaders' preference for "an absence of criticism" that puts their pride, jobs, and funding at risk.[150] In 2009, Saturday Night Live hit on that

---

Kim & Simon Denyer, *South Korea Is Doing 10,000 Coronavirus Tests a Day. The U.S. Is Struggling for Even a Small Fraction of That.*, WASH. POST (Mar. 13, 2020), https://www.washingtonpost.com/world/asia_pacific/coronavirus-test-kits-south-korea-us/2020/03/13/007f14fc-64a1-11ea-8a8e-5c5336b32760_story.html [https://perma.cc/WR8L-XCQB] (referencing prior outbreaks).

146. BD. OF GOVERNORS OF THE FED. RSRV. SYS., 2020 SUPERVISORY SCENARIOS FOR ANNUAL STRESS TESTS REQUIRED UNDER THE DODD-FRANK ACT STRESS TESTING RULES AND THE CAPITAL PLAN RULE 3 (Feb. 2020), https://www.federalreserve.gov/newsevents/pressreleases/files/bcreg20200206a1.pdf [https://perma.cc/V2WU-JWPU].

147. *Id.* at 4; Harriet Torry, *U.S. Economy Contracted at Record Rate Last Quarter; Jobless Claims Rise to 1.43 Million,* WALL ST. J., https://www.wsj.com/articles/us-economy-gdp-report-second-quarter-coronavirus-11596061406 (last updated July 30, 2020, 3:21 PM) [https://perma.cc/U7KW-WP8M].

148. *See supra* Section I.A (summarizing the psychology of crises).

149. Karma Allen, *Dimon: Another Crisis Is Inevitable,* CNBC, https://www.cnbc.com/2015/04/08/dimon-defends-jpmorgans-size-breadth.html (last updated Apr. 9, 2015, 9:08 AM) [https://perma.cc/ET2G-MZKQ] (internal quotation marks omitted).

150. JAMES Q. WILSON, BUREAUCRACY: WHAT GOVERNMENT AGENCIES DO AND WHY THEY DO IT 197 (1989) ("The head of a public agency is judged and rewarded on the basis of the *appearance* of success, when success can mean reputation, influence, charm, the absence of criticism, personal

tension and showed that stress tests are not too arcane for reputational damage to financial regulation leaders. Its skit depicted Secretary Geithner explaining that he had decided against using a 1 to 100 grade for stress tests and would instead use a "PASS/PASS system" to be "less judgmental."[151]

Agencies' desires to minimize negative perception may explain why FEMA and HHS designed the Louisiana Hurricane and Crimson Contagion exercises as nonevaluative tests.[152] Even though those agencies largely kept the exercises' existence secret, the hundreds of outside participants made complete confidentiality within the government impossible.[153] Any final report harshly judging agencies risks being leaked to the public or obtained through the Freedom of Information Act[154] and would make it more difficult for agency heads to exaggerate their self-assessments in their annual reports. Even well-intended agency officials might prefer to conduct the exercises in a less evaluative manner, with the hope of learning something without creating political headwinds.[155]

In particular, stress tests tend to lack evaluation that would communicate the big-picture implications of failure—in other words, the full consequences. For instance, neither the FEMA hurricane nor the HHS pandemic simulations told participants how many lives or billions of dollars they could have saved with better judgment or preparation.[156] Rather than an output, harm was a fixed input in the exercises—asking what the CDC would do to combat a rapidly rising

---

ideology, or victory in policy debates."); *see also* Steven P. Croley, *Public Interested Regulation*, 28 FLA. ST. U. L. REV. 7, 11 (2000) ("Congress can increase or decrease an agency's budget, depending in part on Congress' assessment of the agency's performance.").

151. Saturday Night Live, *Geithner Cold Open: Bank Stress Test*, YOUTUBE (Oct. 3, 2013), https://www.youtube.com/watch?v=CERiQe9PbKs [https://perma.cc/KLR4-EQPT] ("[W]e decided that that might unfairly stigmatize banks who scored low on the test because they followed reckless lending practices or were otherwise not good at banking."). Bank failures would potentially reflect poorly on the regulators themselves, which means that a lenient system for banks could shield regulators from scrutiny of the test results—of course, only if that leniency is not itself criticized. For a discussion of such perceptions in bank stress tests, see Baradaran, *supra* note 9, at 1297.

152. *FEMA Simulation Comm. Hearing*, *supra* note 59, at 11. There is no evidence that HHS even produced a final report. Sanger et al., *supra* note 67.

153. *See, e.g.*, *FEMA Simulation Comm. Hearing*, *supra* note 59, at 2 (noting that the Hurricane Pam exercise involved at least three hundred emergency response officials from local, state, and federal governments).

154. *See* 5 U.S.C. § 552(a) (detailing what agencies are required to make public).

155. *Cf.* Kristin N. Johnson, *Macroprudential Regulation: A Sustainable Approach to Regulating Financial Markets*, 2013 U. ILL. L. REV. 881, 901 ("[I]intentional manipulation can undermine the value of these quantitative models.").

156. HHS PANDEMIC SIMULATION DRAFT REP., *supra* note 68; *FEMA Simulation Comm. Hearing*, *supra* note 59.

count of 110 million Americans infected within a few months.[157] With human suffering held constant, the payoff in terms of lives saved for an agency acting on stress test lessons is less clear.

* * *

In sum, even when they participate in stress tests, administrative agencies sometimes fail to take clear steps that could prevent or mitigate catastrophes. Often the failure is not due to any absence of authority or resources but to something entirely in their discretion. Agencies have missed key issues in advance, such as risky mortgages or inadequate virus testing; allowed coordination challenges to slow important decisions; and neglected to create emergency plans.

These case studies of societal challenges from public health, the environment, and finance do not exhaust the law's shortcomings in preparing for potentially acute threats to societal well-being. Stress tests complement other strategies for disaster responses, such as regulatory monitoring. Thus, flawless stress tests would by no means be sufficient by themselves. Stress tests are, however, an underappreciated tool deployed to push an organization toward resilience.

In theory, the pervasiveness of stress tests is a promising development for counteracting psychological tendencies to underprepare for rare events. But as currently utilized, key stress tests are secretive, unimaginative, voluntary, and inconsequential. As a result, agency leaders will often face predictable, time-sensitive decisions for the first time during an actual emergency, with inadequate plans in place.

II. REDESIGNING STRESS TESTS

It would be unrealistic to expect any government to be fully prepared for all future contingencies. However, since stress tests are already part of the playbook, agencies could at least conduct them as effectively as possible. This Part explores four ways to improve the use of stress tests: legal mandates, external oversight, imaginative inputs, and consequential outputs.[158] Since three of these are within administrative discretion, agency heads or the President can

157. HHS PANDEMIC SIMULATION DRAFT REP., *supra* note 68.

158. This list is not exhaustive and aims to sketch initial priorities for a necessarily larger governmental transformation.

significantly reform the state's crisis responsiveness without legislative action.[159]

A caveat is in order. The costs of any change to stress tests explored below must be weighed against the benefits—even if the costs are limited to information management or diversion of current employees' attention from other matters. That cost-benefit analysis is important and the discussion below aims to inform such an analysis. But providing answers to the cost-benefit analysis is not the focus here. Instead, the goal is to unearth the array of potential design improvements for stress tests.

Although these design improvements would need to be subjected to a cost-benefit analysis, many of the ideas discussed below are relatively inexpensive given that stress tests already happen, which means that many design improvements are possible through a reallocation of existing expenditures. Moreover, any increase in stress test costs must be viewed in the context of the already large-scale expenditures devoted to risk management in administrative agencies and private firms.[160] Assuming considerable expenditures for crisis administration will continue, the question of how much to invest in improving stress tests is dependent on comparing indeterminate cost-benefit analyses of the various risk management tools available. To make that comparison rigorously, it would be ideal to understand what stress tests have to offer if designed well.

A. Mandates for Agency and Business Leaders

Despite limitations, conducting stress tests is considered a best practice for organizations involved in high-stakes risks.[161] Some private stakeholders have recognized this value without the law pushing them to do so. For instance, a growing number of businesses hire outside cybersecurity firms to try to breach their defenses through any means possible, including technical hacking, cold-calling employees to obtain

159. On the considerable discretion agencies possess in implementing laws, see, for example, Heckler v. Chaney, 470 U.S. 821, 838 (1985) (finding that the decision not to take enforcement action or institute investigative proceedings is "committed to agency discretion" (internal quotation marks omitted)). *See also* Urska Velikonja, *Accountability for Nonenforcement*, 93 NOTRE DAME L. REV. 1549, 1549 (2018) (discussing Supreme Court precedent that "affords agencies considerable enforcement discretion").

160. *Cf.* Maxwell C. Smith, Anita Ghosh & Catherine E. Kanatas, *Death v. Taxes: Agency Approaches to Setting Safety Goals Using Risk Management in an Evolving Legal Environment*, 26 N.Y.U. ENV'T L.J. 41, 82 (2017) ("[O]ver the past few decades, agencies have heavily invested time and resources in enhancing their risk management.").

161. *Cf.* Weber, *supra* note 9, at 2241 ("Conceptualizing stress testing regulation as a corporate governance matter makes sense.").

passwords, and delivering pizza to infiltrate company premises.[162] These cybersecurity exercises involve employees walking through the steps they would take in response to such assaults.[163]

A recognition of those organizational best practices helps explain lawmakers' statutory insistence on financial regulation simulations, as "enforced, mandatory deliberation on stress and failure might serve as a partial antidote to the tendencies of corporate decision makers, whether individual or group-level, to overuse heuristics in dangerous ways."[164] The availability heuristic and other psychological limits are thus a core justification for imposing stress tests on the private sector.[165]

Government leaders are not immune to such errors. Nor are business leaders in sectors other than banking. Just as banks are not trusted to take adequate precautions on their own, leaders of agencies and other businesses should not have such discretion when the consequences are great and their track records poor.

Therefore, where stress tests of agencies' emergency readiness are beneficial, Congress should mandate them. They should also consider what other types of businesses, such as offshore oil and tech companies, might benefit from similar mandates. Given institutional limits, Congress may want to delegate much of the authority to decide what stress tests should be mandated. A meta-agency might oversee stress tests across agencies.[166] A more embedded model would be to task a staff member or group inside the agency with focusing on stress tests.

Accompanied by the type of design improvements discussed in subsequent sections, mandates would help improve the chances that the extensive resources currently devoted to stress tests are deployed effectively, rather than in ways that insulate agencies from reputational damage.[167] Deciding where stress tests would be appropriate is a difficult task requiring policymakers to weigh

---

162. *See, e.g.*, *The Cyber Security Stress Test*, F-SECURE, https://campaigns.f-secure.com/cyber-security-stress-test/ (last visited Feb. 17, 2022) [https://perma.cc/U7EA-MFUL]; *Cybersecurity Stress-Testing: Don't Stress About Your Company's Safety*, AXIOM CYBER SOLS., https://axiomcyber.com/cybersecurity/cybersecurity-stress-testing-dont-stress-about-your-companys-safety/ (last visited Feb. 17, 2022) [https://perma.cc/75J5-QBJ9].

163. *See generally* sources cited *supra* note 162.

164. Weber, *supra* note 9, at 2242.

165. *See id.* ("[R]egulators should understand their task as involving management-based regulation—a regulatory approach that acts on corporate planning processes . . . .").

166. *See infra* Section II.B.

167. In some contexts, agencies do have some incentives to design stress tests in ways that would increase funding, meaning that they would not want to make it seem as though the agency has everything under control. However, reputational management is a powerful driver of administrative behavior, and self-serving stress tests could always make it seem as though the problem is inadequate resources while avoiding any assumption of responsibility.

alternatives and multiple risks in deciding on the optimal level of precaution.[168] Agencies that both already invest considerable resources in crisis management and have unsatisfactory track records of preparedness are leading candidates for considering such mandates.

Agencies could still go beyond these legislative directives to create additional tests that Congress may not have foreseen. But to avoid the delays that have undermined past agency exercises, Congress should include in any mandate a requirement that the tests occur on a periodic basis, depending on the agency and nature of the potential crisis.

Each test would seek to answer different questions, mostly chosen by the agency, Congress, and other external stakeholders.[169] Is the Federal Reserve prepared to handle financial instability driven by the shadow banking system even though it cannot directly regulate key actors such as Venmo, Credit Karma, Mint, and cryptocurrency firms? Is FEMA ready to coordinate evacuation and cleanup in the face of an unprecedented storm or toxic spill carried nationwide through the air? How would the CDC and FDA accelerate virus screening for the next pandemic?

To explore these questions, it would sometimes be necessary to design stress tests at the cross-agency and department level, as was the case for Crimson Contagion. Yet the key sites of decisionmaking are often in agencies within agencies.[170] For example, HHS oversees eleven agencies, including visible ones such as the CDC, National Institute of Health ("NIH"), and FDA, as well as the lesser known Office of the Assistant Secretary for Preparedness and Response ("ASPR").[171] The

168. *Cf.* Jonathan B. Wiener, *Precaution in a Multirisk World*, *in* HUMAN AND ECOLOGICAL RISK ASSESSMENT: THEORY AND PRACTICE 1509, 1526 (Dennis J. Paustenbach ed., 2002) (discussing the challenge of managing multiple risks and observing "the goal should be not maximum precaution but an 'optimal precaution' that addresses both the risks of inaction and the risks of action").

169. *See infra* Section II.B (discussing external oversight).

170. The Administrative Procedure Act defines agencies to include larger authorities that have agencies within them. 5 U.S.C. § 551(1) (" '[A]gency' means each authority of the Government of the United States, whether or not it is within or subject to review by another agency . . . .").

171. *HHS Agencies & Offices*, U.S. DEP'T OF HEALTH & HUM. SERVS., https://www.hhs.gov/about/agencies/hhs-agencies-and-offices/index.html (last updated Oct. 27, 2015) [https://perma.cc/W9G8-HYZ4]. The ASPR oversees the Biomedical Advanced Research and Development Authority, whose mission includes developing, manufacturing, and distributing vaccines and other medical countermeasures in the event of a public health emergency. *BARDA Strategic Plan 2011-2016*, U.S. DEP'T OF HEALTH & HUM. SERVS. 5 (2011), https://www.phe.gov/about/barda/Documents/barda-strategic-plan.pdf [https://perma.cc/HE97-YKXJ] (listing other goals that include responding to chemical, biological, radiation, and nuclear attacks).

CDC, NIH, and FDA are among the largest federal agencies, each with over ten thousand employees.[172]

Each crucial senior management team must know what it needs to do to prepare for a crisis.[173] Consequently, a single stress test of the sprawling HHS alone cannot adequately prepare so many diverse entities for a pandemic. Moreover, since the law often mandates that businesses assume regulatory roles,[174] in such contexts those firms may be conceptualized as analogous to agencies within agencies for purposes of possibly mandating stress tests.

Not all agencies will need legal mandates. Some federal entities, such as the military, may already have proper incentives to regularly run robust simulations.[175] For others with less crisis-oriented missions, stress tests seem potentially inappropriate. It is not immediately clear, for instance, how the Equal Employment Opportunity Commission could stress test its ability to prosecute workplace discrimination, or why the Federal Trade Commission ("FTC") would need to adopt the practice for its consumer protection, privacy, and antitrust enforcement.[176]

On the other hand, stress tests—or something like them—could be imagined for the FTC if stress tests were not defined as only appropriate for threats that arrive suddenly or impose harm quickly. By relaxing this limiting principle of speed, it becomes more possible to envision simulations like those currently used by crisis agencies also helping the FTC get ahead of big problems within its regulatory mandate. According to some scholars, the agency failed to get ahead of widespread privacy violations, and "[l]oss of control over personal information creates a variety of near-term and longer-term risks that

---

172. *Federal Workforce Data*, U.S. OFF. OF PERS. MGMT., https://www.fedscope.opm.gov/employment.asp (last visited Feb. 17, 2022) [https://perma.cc/SH5Q-CPKX] (providing official employment figures for federal agencies and classifying the FDA and CDC as "large").

173. *Cf.* Thomas H. Stanton, *The Growing Movement for Enterprise Risk Management in Government: The United States Begins To Catch Up*, 37 ASIA PAC. J. PUB. ADMIN. 182, 185 (2015) ("In the US federal government, the agency rather than the department is often the optimal level for [enterprise risk management].").

174. *See, e.g.*, Rory Van Loo, *The New Gatekeepers: Private Firms as Public Enforcers*, 106 VA. L. REV. 467 (2020) (using case studies from the oil, technology, pharmaceutical, and banking industries to show how firms are increasingly conscripted to enforce the law against other firms).

175. Academics tend to offer the military as an example of simulation best practices. *See* Baradaran, *supra* note 9, at 1319 (arguing that financial regulators should adopt wargame practices from the military); James A. "Sandy" Winnefeld, Jr., Christopher Kirchhoff & David M. Upton, *Cybersecurity's Human Factor: Lessons from the Pentagon*, HARV. BUS. REV., Sept. 2015, at 215 (summarizing the military's widespread use of war games). Nonetheless, an institution whose performance sets the standard is not beyond opportunity for improvement.

176. On the FTC's authority in these areas, see 15 U.S.C. §§ 41-58.

are difficult to understand and value."[177] Most prominently, the sharing of social network data without permission allowed third parties such as Cambridge Analytica to attempt to steer the 2016 election toward then-candidate Donald Trump.[178] According to other critics, the FTC has also failed to grasp the implications of allowing the five biggest tech firms to purchase over four hundred companies, including Facebook buying Instagram and WhatsApp, and Google acquiring Double-Click, YouTube, and Waze.[179] The failure to block harmful mergers, combined with price increases due to manipulation or deception, could contribute to economic inequality on a large scale.[180]

Thus, it is not inconceivable that role-playing a doomsday simulation related to privacy, election misinformation, or economic inequality could help the FTC understand the full societal implications of unfettered data access and unchecked acquisitions in the tech sector. The FTC could thereby assess what actions now and later would be appropriate under varying sets of assumptions. Or the FTC might require stress tests of certain companies, such as platforms, to ensure that those firms are adequately prepared for potential dire consequences of their operations.[181]

Although such simulations would stretch the concept of stress tests beyond their common understanding, what to call such exercises is less important than determining whether they would improve governance. This Article seeks to encourage conversations about the scope and role of stress tests—and their cognates—rather than to provide an exhaustive list or typology of agencies' suitability. The right answer will surely be context specific, and in agencies such as the FTC, it may look less like a stress test and more like threat modeling. Yet it would be premature to dismiss outright the possibility of tailoring stress tests for noncrisis agencies.

---

177. Cohen, *supra* note 17, at 377.

178. *See, e.g.*, Philip M. Napoli, *What If More Speech Is No Longer the Solution? First Amendment Theory Meets Fake News and the Filter Bubble*, 70 FED. COMMC'NS L.J. 55, 75–76 (2018) (exploring the free-speech implications of data sharing).

179. *See, e.g.*, TIM WU, THE CURSE OF BIGNESS: ANTITRUST IN THE NEW GILDED AGE 132–33 (2018) (advocating the dissolution of big tech companies); C. Scott Hemphill, *Disruptive Incumbents: Platform Competition in an Age of Machine Learning*, 119 COLUM. L. REV. 1973, 1987–88 (2019) (highlighting the competitive threat that Instagram posed to Facebook before the acquisition); Astead W. Herndon, *Elizabeth Warren Proposes Breaking up Tech Giants Like Amazon and Facebook*, N.Y. TIMES (Mar. 8, 2019), https://www.nytimes.com/2019/03/08/us/politics/elizabeth-warren-amazon.html [https://perma.cc/9KXE-QG92] (summarizing the plans by Elizabeth Warren, Bernie Sanders, and others to break up big tech companies).

180. *See* Rory Van Loo, *Broadening Consumer Law: Competition, Protection, and Distribution*, 95 NOTRE DAME L. Rev. 211 (2019) (positing that antitrust and consumer protection could combine to play a significant role in lowering income inequality).

181. This topic is explored further *infra* Part III.

Ensuring that mandated stress tests result in action will be difficult, and the following sections turn to that task. But without a mandate, valuable stress tests may not occur at all, or may happen too infrequently or inadequately to be useful.

B. External Oversight

Simply conducting stress tests is not enough. The law would ideally also increase the chances that agencies act on the lessons learned from stress tests when it would be overall beneficial to do so. One strategy is to involve external stakeholders in all phases. This involvement can combine visibility, input, and supervision.

To increase accountability, mandated stress tests should involve some external actor outside the agency. At a minimum, agencies could brief Congress and the White House on the results of the most important simulations and provide subsequent updates on progress towards resolving shortcomings. But Congress has limited attention and resources and cannot engage in the design details or even summary results from all agency stress tests. Additionally, underperforming executive agencies can reflect poorly on the president. Consequently, the White House also has incentives to produce self-serving evaluations. Nonetheless, requiring that key stress test results be sent to the president makes sense in some areas, and would align other agencies with the military's tendency to update the president on significant war game results.[182]

Governmental actors, such as the Government Accountability Office ("GAO"), General Services Administration's Office of Evaluation Sciences, National Academy of Sciences, Inspector General, or the Congressional Budget Office could provide additional support and oversight. One of the most straightforward existing homes for stress-test oversight is the GAO, which has already conducted studies of different agencies' enterprise risk management.[183] A new bureau, whether standalone or within an existing entity, could also focus on stress tests and risk management oversight of other agencies.

---

182. *See, e.g.*, Winnefeld et al., *supra* note 175, at 94 (describing the military's general practices for security audits).

183. U.S. GOV'T ACCOUNTABILITY OFF., GAO-17-63, ENTERPRISE RISK MANAGEMENT: SELECTED AGENCIES' EXPERIENCES ILLUSTRATE GOOD PRACTICES IN MANAGING RISK 4 (2016) (interviewing twenty-four agencies on their established enterprise risk management policies). The GAO's role would need to abide by separation of power constraints. *See, e.g.*, Bowsher v. Synar, 478 U.S. 714, 733 (1986) (ruling on allowable executive powers).

Concerns about managing widespread alarm speak to a well-known gap between perceptions by the government and public.[184] That challenge "puts the regulators in the awkward position of defending expert risk perceptions as more valid or rational than the public's."[185] The greater society's "confidence/trust in institutions and elites," the more likely the populace will defer to established institutions and laws.[186]

However, while some details must be kept confidential, particularly sensitive information about private actors or intelligence matters, those justifications for secrecy are less relevant to stress testing public agencies. Another risk is that too much transparency with stress tests could pressure agency leaders to prepare excessively in order to shield themselves from later being criticized. Such risks are worth factoring into the implementation of transparency into stress tests.

Those risks must be weighed against the possible benefits of public transparency allowing society to advocate for its interests and to hold leaders accountable for their failures.[187] Publicly available stress tests would allow diverse stakeholders to assess the results, thus discouraging or weakening self-serving proclamations of success.[188] Indeed, unnecessary secrecy can make the public more prone to panic and crises harder to manage because people lack faith in leaders and are therefore less likely to heed public warnings.[189] Moreover, in a real crisis, agencies would need to deal with heightened public scrutiny, so transparency could create more realistic simulations by making officials manage the public.

Thus, agencies would ideally default to releasing the results of stress tests publicly and assuaging worries by taking appropriate action or explaining their nonaction. That approach is preferable to agencies

184. *See* Lori S. Bennear, *Economic Analysis, Risk Regulation, and the Dynamics of Policy Regret*, *in* POLICY SHOCK, *supra* note 39, at 43, 53 ("[P]ublic perceptions of risk often differ . . . from expert perceptions of risk . . . .").

185. *Id.* at 54.

186. Edward J. Balleisen, Lori S. Bennear, Kimberly D. Krawiec & Jonathan B. Wiener, *Recalibrating Risk: Crises, Learning, and Regulatory Change, in* POLICY SHOCK, *supra* note 39, at 540, 549.

187. *See, e.g.*, Rebecca Bratspies & Sarah Lamdan, *Taking a Page from the FDA's Prescription Medicine Information Rules: Reimagining Environmental Information for Climate Change*, 40 U. ARK. LITTLE ROCK L. REV. 573, 583 (2018) (discussing transparency in both the health and environmental contexts and arguing that "public access to data helps communities make informed decisions about the chemical hazards in their midst").

188. *See supra* Section II.A (discussing the importance of external stakeholders).

189. ROBERT GALFORD & ANNE SEIBOLD DRAPEAU, THE TRUSTED LEADER 210–11 (2002) (describing how organizations and individuals with low levels of trust lack the resources to fight threats because "[t]hey lack the ability to engage, or bring out the best in their people").

promoting calmness with silence, inaction, and misinformation. Avoiding panic is more relevant during emergencies than between them—when most stress tests would occur. During those lulls, alarming simulation conclusions may be necessary to prompt authorities to adequately prepare for a future contingency.[190]

Although the primary goal would be organizational change, stress tests can also feed into policymaking. Based on results, agencies could reallocate existing resources or write relevant rules. In some instances, agencies would need to ask Congress for either new legislation or funding to address identified shortcomings. Providing transparency and oversight for stress tests thus not only reflects good governance principles, but also can help address irrational inaction.

C. Imaginative Scenarios

Stress tests should be designed by choosing scenarios from a comprehensive set of predictions. A robust literature has explored ways for policymaking to incorporate the future in a more effective manner.[191] Comprehensive predictions will sometimes mean imagining unlikely scenarios. Imagination can help envision which disasters may arise and predict what challenges those disasters will pose once they do. Since people exhibit natural psychological resistance to thinking about negative topics and inherent bias toward underestimating unlikely future events,[192] the law has a role to play in forcing imagination.

Financial regulation scholars are aware of the lack of creativity in bank stress tests and the resulting need "to normalize (rather than problematize) unexpected events that might be weak-signal harbingers of future catastrophe."[193] If regulators had paid more attention to scholarly warnings about mortgage markets in the years leading up to the 2008 financial crisis,[194] or analyzed their data more effectively, they would not have been so blindsided.

---

190. The federal government has sometimes begun to act on a recognition of inadequate crisis preparation in the wake of a near miss, only to eventually lose interest in that project. For example, the federal government identified a need for a larger stockpile of ventilators thirteen years before the coronavirus pandemic, and after the contracts fell through, is still waiting to get any ventilators delivered. Kulish et al., *supra* note 77. Having a regular stress test could continue to push those authorities toward action as the near miss fades from memory.

191. *See, e.g.*, Jonathan B. Wiener, *Best Cass Scenario*, 43 TULSA L. REV. 933, 943 (2008) (situating the "venerable literature on the development of scenarios" in the context of explaining "why people envision the future as they do, and how to do better").

192. *Supra* Section I.A.1.

193. Weber, *supra* note 9, at 2242.

194. *See* Engel & McCoy, *supra* note 133, at 1366 (warning of systemic failures in mortgage markets in 2002).

That creativity deficit also affects other regulatory spheres. For instance, after terrorists took the unprecedented steps of hijacking commercial airplanes and crashing them into the World Trade Center in New York and the Pentagon in Washington, D.C., the congressionally appointed 9/11 Commission recommended "routinizing, even bureaucratizing, the exercise of imagination" to lessen the chance of being caught unaware.[195] It bears emphasis that imagination goes not only to the main trigger, such as a worldwide pandemic, but also to subsequent details like new variants of the virus that make vaccines less effective.

Agency heads can encourage imagination through their leadership. But there are also ways to institutionalize it. As one example, administrative agencies could create a forecaster, futurist, or strategist position. Regardless of the name, the job description would emphasize creating and analyzing contingencies. These strategists' predictions would serve as the bases for simulations—whether for the agency itself or the banks, oil companies, and other private entities it oversees. The strategists would then promote adequate attention towards those contingencies and advocate for agency action to address the lessons learned.

The private sector has created leadership roles for similar purposes. Ford's Chief Futurist alerts the company's auto designers to looming changes like water scarcity, even though she does not have an automotive background.[196] As another example, Citigroup recently hired its first Global Head of Emerging Platforms and Services, quickly dubbed the "Citi Futurist," to serve as a "visionary strategist" and develop banking products for mobile apps and wearable devices.[197]

Irrespective of the title and specific responsibilities, for these individuals to have meaningful institutional influence, they must sit high in the organizational hierarchy.[198] Agencies must also allocate

195. NAT'L COMM'N ON TERRORIST ATTACKS UPON THE U.S., THE 9/11 COMMISSION REPORT 344 (2004), http://www.9-11commission.gov/report/911Report.pdf [https://perma.cc/5KF2-4BRT].

196. Christine Bader, *Beyond the Chief Sustainability Officer*, ATLANTIC (May 20, 2015), https://www.theatlantic.com/business/archive/2015/05/other-corporate-social-responsibility-jobs/393677/ [https://perma.cc/3MZ4-U7Z6].

197. Penny Crosman, *Citi Futurist Envisions the Part Wearables Will Play in Banking*, AM. BANKER (Jan. 16, 2015, 2:43 PM), https://www.americanbanker.com/news/citi-futurist-envisions-the-part-wearables-will-play-in-banking [https://perma.cc/3MCY-RNDR] (covering Citigroup's new position).

198. Ideally both the department and the sub-agencies would have such roles—for example, both HHS and the CDC. For further examinations of involving third parties in the regulatory process, see Brett McDonnell & Daniel Schwarcz, *Regulatory Contrarians*, 89 N.C. L. REV. 1629, 1630 (2011); and Robert F. Weber, *New Governance, Financial Regulation, and Challenges to Legitimacy: The Example of the Internal Models Approach to Capital Adequacy Regulation*, 62 ADMIN. L. REV. 783, 854 (2010).

sufficient resources to these strategists so they can rigorously research scenarios, model ideas, and drive necessary organizational transformations. Research is an essential component of comprehensive and imaginative predictions because history, other countries' experiences, and emerging current events indicate what is possible. Since their analyses should include not only ethical and qualitative considerations but also economic implications, they would ideally have a team of cross-functional experts. Ford's Chief Futurist, for instance, leads an interdisciplinary "trends team" to study topics like how Japan and the cell phone industry are responding to aging populations.[199]

Some agencies have already received prescient insights from a related approach: hiring future-oriented consulting firms. Chevron, Microsoft, Sony, the Department of Defense, the Department of State, and others have commissioned The Future Today Institute, which "researches, models and prototypes future risk and opportunity."[200] In 2014, HHS, the CDC, and other public health agencies engaged the Institute for Alternative Futures in crafting a report imagining the public health system in 2030.[201] To illustrate how imaginative these groups can be, the institute's staff, who have titles such as "Senior Futurist," produced four scenarios and held a national workshop in which agency participants rated each scenario's likelihood and developed relevant strategies.[202] The scenario rated as second most likely—having a fifty percent chance of occurring—included public health agencies mishandling a 2020 flu pandemic so badly that a hostile political climate subsequently led to health care funding cuts and greater health inequality.[203]

Similarly, FEMA hired a private contractor, Innovative Emergency Management, that simulated "with eerie accuracy the all-too-real problems of Katrina."[204] The contractor wrote the scenario with the storm leaving 1.1 million people homeless, producing twenty inches

199. Bader, *supra* note 196.

200. *About The Future Today Institute*, FUTURE TODAY INST., https://futuretodayinstitute.com/about/ (last visited Feb. 17, 2022) [https://perma.cc/285L-L6QB].

201. INST. FOR ALT. FUTURES, PUBLIC HEALTH 2030: A SCENARIO EXPLORATION 54 (2014), https://wayback.archive-it.org/13466/20200205000421/https://altfutures.org/wp-content/uploads/2016/04/IAF-PublicHealth2030Scenarios.pdf [https://perma.cc/D77Q-XVYY].

202. *Id.* at 39 (explaining that workshop participants "stepped into" the scenarios to develop strategies).

203. *Id.* at 12–17, 39. The most likely scenario, rated as sixty-two percent likely to occur, imagined that "public health agencies and health care slowly advance their capabilities" but considerable "variations in technological capabilities, funding, and approaches to prevention—along with a continuous rise in health care costs—significantly limit public health gains." *Id.* at 6, 39.

204. *FEMA Simulation Comm. Hearing*, *supra* note 59, at 2, 15.

of rain, and flooding or shutting down fifty-seven chemical plants.[205] Just one year later, real-life Katrina left one million people homeless, produced eighteen inches of rain, and flooded or shut down over fifty plants.[206] More importantly, the forecasters identified the array of issues that government officials would face.[207] However, FEMA reduced funding after the multi-month simulation had begun, eliminating what became one of the most accurate hypotheticals requiring a solution: How should authorities respond to storm waters breaching the levees?[208]

It is possible that an institutionalized futurist position could have made better use of the external predictions produced. Ideally, whether an internal office or external consultant provides the imaginary vision, there would be some transparency into the proposed scenarios to further the essential job of ensuring crucial tests are fully carried out.

It would be unrealistic to expect agencies, their dedicated prognosticators, or outside consultants to always be right. Indeed, it is surely more likely than not that out-of-the-box risks identified will not unfold exactly as predicted. It is nonetheless worth starting the stress test process with as full a range of threats and scenarios as possible. Although some agencies have, in an ad hoc manner, deployed imaginative thinking, there is little evidence that regulators are systematically integrating such analyses. Embedding inherently visionary posts high in the administrative power structure would offer a means to construct a less retrospective and more prospective administrative state.

D. Consequential Outputs

To make it more likely that agency leaders make decisions about whether to act based on a comprehensive cost-benefit analysis, stress tests should communicate the full consequences of not improving organizational performance. A threshold step is to candidly evaluate the organization. Most agency-focused simulations provide participants with learning experiences and perhaps produce a list of identified issues and self-generated strategies.[209] In other words, they are not much

205. *Id.* at 17.

206. *Id.*

207. *Id.*

208. *Id.* at 2.

209. *See, e.g.*, INST. FOR ALT. FUTURES, *supra* note 201, at 39 (allowing the participants to determine takeaways); HHS PANDEMIC SIMULATION DRAFT REP., *supra* note 68, at 2–5 (providing

more evaluative than the Saturday Night Live "PASS/PASS" system for bank stress tests.[210] Those parodied tests give each bank quantitative performance feedback and a resiliency rating.[211] Administrative stress tests should also provide each agency with frank feedback and perhaps something like a grade.

Another component of communicating consequences is translating results into social implications. Most immediately, the test should make it clear if the agency's inadequate planning heightened the risk of a deep recession, destructive fire, or other major harm. The next level of output is projecting the harm produced by those higher risks, such as the tens of thousands of expected *additional* lives lost due to the agency's inadequate response to a simulated influenza.

The more comprehensive the understanding of harm the better. For many calamities, that means physical, social, and economic devastation. If a public health stress test were to only report the implications of a pandemic in terms of lives lost, it would underestimate the consequences by omitting the associated financial toll and exacerbated inequality.[212]

Comprehensive harm outputs would help focus participants' minds on the stakes, thereby increasing the chances they will remember and learn from the experience.[213] Such outputs also serve a political purpose. Making the economic case for preventing devastation can prompt action.[214] Research has established that policies benefitting high-income households are far more likely to materialize.[215] Thus, estimating the economic implications of crisis prevention makes passing corresponding policies more likely because influential stakeholders would better understand the considerable wealth they may stand to lose without public investments in emergency management. Transparency alone may be ineffective unless stress tests adequately convey the complete stakes of performing better to rouse the media, public, and other stakeholders.

---

nonevaluative top-level summaries); *FEMA Simulation Comm. Hearing*, *supra* note 59, at 2–3 (describing a largely non-evaluative simulation).

210. *See supra* note 151 and accompanying text.

211. HAL S. SCOTT, CONNECTEDNESS AND CONTAGION: PROTECTING THE FINANCIAL SYSTEM FROM PANICS 176–77 (2016).

212. *See supra* Section I.B.2.

213. *See supra* Section I.A.1 (discussing the availability heuristic and tendency to increase risk-taking following near misses).

214. *See supra* Section I.B.3 (discussing stress tests for financial systems).

215. *See generally* MARTIN GILENS, AFFLUENCE AND INFLUENCE: ECONOMIC INEQUALITY AND POLITICAL POWER IN AMERICA (2012).

There is no single agreed upon set of parameters for analyzing crisis management.[216] Inevitably, comprehensive outputs would involve uncertainty. Regulators must therefore make outputs' incertitude clear. Failing to do so would create an illusion of precision.[217] To lessen such misperceptions, regulators could give ranges of estimates, clearly disclose limitations, and label figures as worst-case, most likely, or another descriptor.[218]

It would be a mistake, however, to omit comprehensive outputs because of uncertainty. Similar indeterminacy plagues most areas of governance—including estimates of how much revenue a tax policy will generate,[219] GDP forecasts,[220] and the likelihood that a merger of two large firms creates monopoly power.[221] Even the main standard for civil trials, "preponderance of the evidence," acknowledges that meaningful decisions must be made with some level of uncertainty since it allows a party to win by establishing a likelihood not far above fifty percent.[222]

The more appropriate question is whether an estimate of an exercise's social implications, regardless of its inability to predict flawlessly, would improve governance. Without any rigorously researched estimate, the mind is left to its own optimism biases and other psychological limitations—which include poor probability construction.[223] For instance, "the "vast majority of people believe that they are less likely than others to be subject to automobile accidents, infection from AIDS, heart attacks, asthma, and many other health risks," even though statistically speaking they cannot all be less likely than most others.[224] Those individual tendencies toward excess risk-taking are even more complicated in an organizational setting and influence "the role of the regulatory state, especially in the area of

216. *See* Bennear, *supra* note 184, at 49 (describing options such as "(1) the utility function, (2) discount rate, (3) risk aversion, and (4) subjective probabilities").

217. *Cf.* Michael J. Graetz, *Paint-by-Numbers Tax Lawmaking*, 95 COLUM. L. REV. 609, 613–14 (1995) (commenting that lawmakers suffer from an illusion of precision about the accuracy of tax law revenue distribution estimates).

218. *Id.*

219. *See id.* at 625–26 (discussing tax indeterminacy).

220. *See generally* YAIR LISTOKIN, LAW AND MACROECONOMICS: LEGAL REMEDIES TO RECESSIONS 3–4 (2019) (explaining the quality of macroeconomics).

221. Rebecca Haw, *Adversarial Economics in Antitrust Litigation: Losing Academic Consensus in the Battle of the Experts*, 106 NW. U. L. REV. 1261, 1300 (2012) (describing how economic experts partaking in litigation fall within a spectrum of opinions in their field of study).

222. Kevin M. Clermont, *Trial by Traditional Probability, Relative Plausibility, or Belief Function?*, 66 CASE W. RSRV. L. REV. 353, 355 (2015) (summarizing the uncertainty involved in determining the preponderance-of-the-evidence standard and how juries struggle to conceptualize proof in terms of percentages).

223. *See, e.g.*, Christine Jolls, Cass R. Sunstein & Richard Thaler, *A Behavioral Approach to Law and Economics*, 50 STAN. L. REV. 1471, 1477 (1998).

224. Cass R. Sunstein, *Behavioral Analysis of Law*, 64 U. CHI. L. REV. 1175, 1183 (1997).

dangers to life and health."[225] Understanding should thus be viewed on a spectrum. Comprehensive outputs can move people from the impressionistic toward the more empirically informed end of that spectrum.

Despite the inevitable indeterminacy of output estimates, sometimes preparedness can be drastically improved at minimal cost. That is especially true when part of the problem is organizational, such as delegating authority for rapidly approving aid relief in the event of a hurricane striking New Orleans.[226] Most importantly, pressure spurs the otherwise complacent mind to find solutions.[227] Stress tests with comprehensive and evaluative outputs would provide such pressure, particularly on leaders.

* * *

A consistent critique in the legal literature is the administrative state's inertia in response to a technologically, scientifically, and financially fast-changing world.[228] Unlike businesses, which face elimination if they fail, administrative agencies can persist in a state of underperformance with few consequences. Administrative law responds by involving Congress, the president, courts, and the public through annual reports, judicial review, funding allocation, and notice-and-comment rulemaking.[229] Yet those traditional accountability mechanisms are of limited use in prompting agency leaders to rectify organizational psychology challenges such as availability heuristics.

---

225. *Id.*

226. *See supra* Section I.B.2. (discussing extreme weather simulations).

227. *See generally* Shih-Wen Hsiao, Ming-Feng Wang & Chien-Wie Chen, *Time Pressure and Creativity in Industrial Design*, 27 INT'L J. TECH. & DESIGN EDUC. 271 (2017) (finding that a lack of pressure leads to apathy, an appropriate amount of pressure promotes creativity, and excessive pressure results in frustration).

228. *See, e.g.*, ROSEMARY O'LEARY, ROBERT F. DURANT, DANIEL J. FIORINO & PAUL S. WEILAND, MANAGING FOR THE ENVIRONMENT: UNDERSTANDING THE LEGAL, ORGANIZATIONAL, AND POLICY CHALLENGES 332 (1999) (discussing regulators' inability to keep up because "[m]ost regulated industries today are more dynamic and complex than they were thirty years ago"). The sophistication gap facing financial regulators, for example, is pervasive across agencies scrambling to adapt to increasingly complex and digital markets. Scholars have already offered solutions to many of these challenges. *See generally* M. Todd Henderson & Frederick Tung, *Pay for Regulator Performance*, 85 S. CAL. L. REV. 1003 (2012) (proposing a new compensation structure for regulators that awards them for their results); Tom C.W. Lin, *The New Investor*, 60 UCLA L. REV. 678, 682–84 (2013) (proposing principles for managing an increasingly technological financial sector).

229. *See* 5 U.S.C. §§ 553, 702, 704 ("[F]inal agency action for which there is no other adequate remedy in a court [is] subject to judicial review."); Croley, *supra* note 150, at 12 (summarizing oversight mechanisms). Courts have emphasized that judicial review should not extend to a monitoring role over agencies. Allen v. Wright, 468 U.S. 737, 759–60 (1984) (stating that lawsuits challenging "particular programs agencies establish to carry out their legal obligations . . . are rarely if ever appropriate for federal-court adjudication"); Laird v. Tatum, 408 U.S. 1, 15 (1972) ("[Acting] as virtually continuing monitors of the wisdom and soundness of Executive action . . . is not the role of the judiciary . . . .").

Mandatory, transparent, imaginative, and comprehensive stress tests should enter the conversation about how to fill that gap in the toolkit, and thereby may have a considerably larger role to play in administrative law. Implementing these principles is important but will prove challenging. It is not possible to finalize here all of the details of operationalizing stress tests—and undertaking the cost-benefit analysis to ensure that the changes are worth the resource investment.[230] Nonetheless, with these design features as a starting point, stress tests could help address the legitimacy crisis facing administrative agencies by increasing the accountability and effectiveness of delegated authority.[231]

At a minimum, an interdisciplinary analysis of simulation failures and their social, health, economic, and environmental implications should push agencies to implement lower-cost reforms to minimize societal risk. Ultimately, the goal is to improve the level of clarity that stress tests provide about the trade-offs involved in action and nonaction. Ideally, the nation could effectively use simulations to learn valuable lessons in advance of calamities, rather than waiting for them to arrive.

III. APPLYING STRESS TESTS

Designing administrative stress test regimes will require considerable attention to context-specific details. The written materials for a single stress test exercise typically fill hundreds of pages.[232] Consequently, a comprehensive treatment of whether and how any given context merits stress tests is beyond the scope of this Article. Nonetheless, this Part considers how improved stress tests might accelerate change within the administrative frameworks surrounding pandemics, environmental disasters, financial crises, and new private-

230. *See generally* Robin Kundis Craig & J.B. Ruhl, *Designing Administrative Law for Adaptive Management*, 67 VAND. L. REV. 1 (2014) (proposing ways for the law to encourage agencies to respond quickly and adaptively); Dominic D. P. Johnson, Rose McDermott, Emily S. Barrett, Jonathan Cowden, Richard Wrangham, Matthew H. McIntyre & Stephen Peter Rosen, *Overconfidence in Wargames: Experimental Evidence on Expectations, Aggression, Gender and Testosterone*, 273 PROC. ROYAL SOC'Y B 2513 (2006) (discussing how participants in close-call simulations can be overconfident in their ability to emerge victorious in the face of future adverse scenarios).

231. *See, e.g.*, JOHN HART ELY, DEMOCRACY AND DISTRUST: A THEORY OF JUDICIAL REVIEW 133 (1980) ("That legislators often find it convenient to escape accountability is precisely the reason *for* a non-delegation doctrine."). *See generally* DANIEL R. ERNST, TOCQUEVILLE'S NIGHTMARE: THE ADMINISTRATIVE STATE EMERGES IN AMERICA, 1900–1940 (2014) (providing an overview of administrative legitimacy challenges).

232. *See, e.g.*, HHS PANDEMIC SIMULATION DRAFT REP., *supra* note 68 (demonstrating the length involved).

sector contexts. In analyzing the preparation for past crises, there is always a risk of inaccurately concluding in hindsight that some past mistake could have been avoided when the decisions were in fact reasonable given the information available at the time.

The purpose below is neither to pass final judgment on past stress tests nor to propose particular stress tests. Instead, the immediate task is illuminating the path forward in a more concrete manner than the broad principles offered above. The larger goal is to start a conversation about how to improve what has become a major administrative undertaking.

A. Pandemics

Currently, discretionary public health resources need to focus on managing the COVID-19 response. After this virus is no longer a crisis, authorities might deploy stress tests to prepare for the next health threat. One way to accomplish that goal would be for Congress to require regular public health stress tests, at least every two or three years.[233] For such exercises to be successful, it would be important for HHS and its relevant sub-agencies—most notably the CDC and the Office of the Assistant Secretary for Preparedness and Response—to be required to participate. Legislation would also ideally mandate the involvement of the directors of these and other relevant agencies, such as the FDA and NIH.

As one of many public health threats that could be examined, consider how the Crimson Contagion stress test might have been improved. One of the most notable omissions from the simulation was any consideration of virus testing. If virus testing had been included as an input, it would have revealed the glaring weaknesses discussed above, such as the excessively slow process for approving screening centers and scientific testing methods.[234] HHS's failure to include virus testing as one of the considerations in Crimson Contagion demonstrates a fatal shortcoming in determining key issues to evaluate in public health simulations.

We cannot simply dismiss the lack of testing as something that is obvious now but was impossible to predict. From an institutional perspective, if HHS had an effective futurist-led team designing its stress tests, it would have surveyed past critical issues faced elsewhere

233. The frequency of the stress tests should reflect the magnitude of the threat. There is no one right answer in terms of whether smaller stress tests of individual agencies or issues, such as vaccine distribution, would be preferable.

234. *See supra* Section I.B.1.

when generating as comprehensive a list of scenarios as possible.[235] That research would have surely included South Korea's experience with a 2015 outbreak of Middle East Respiratory Syndrome ("MERS"), in which the country's inadequate testing proved deadly.[236]

As a result of that lesson, South Korean health authorities subsequently implemented a collaborative system in which regulators and researchers could respond quickly to supply new testing kits.[237] The country also passed legislation "providing for the immediate approval of testing systems in the event of a health crisis."[238] Those legislative and regulatory responses produced results. For instance, South Korea had its first confirmed COVID-19 death the same week as the United States did.[239] Within seven days, South Korea tested more people per *day*—ten thousand—than the United States tested in the entire *month* after its first death.[240]

That disparity is unsurprising given that early in the pandemic often only a single testing site was available even in populous states, such as Massachusetts.[241] That state's two world-renowned hospitals, Brigham & Women's and Mass General, took weeks to begin testing due to bureaucratic hurdles. In the meantime, the only option was a small CDC site.[242] If it had included the South Korea experience as part of the design, Crimson Contagion would have probed U.S. screening capacity. Following the simulation, the participants could have advocated for the CDC to improve its virus testing and the FDA to adopt a streamlined pandemic screening approval process for independent labs.

Another path for increasing the chances that the Crimson Contagion exercise included screening would have been external oversight of the simulation's design—such as by an independent academic advisory board. Among academic researchers in epidemiology, screening was a well-known determinant of pandemic response success.[243]

---

235. *See supra* Section II.C.

236. Kim & Denyer, *supra* note 145.

237. *Id.*

238. Victor Cha, *South Korea Offers a Lesson in Best Practices*, FOREIGN AFFS. (Apr. 10, 2020), https://www.foreignaffairs.com/articles/united-states/2020-04-10/south-korea-offers-lesson-best-practices [https://perma.cc/49MA-QQNV].

239. Sheri Fink & Mike Baker, *'It's Just Everywhere Already': How Delays in Testing Set Back the U.S. Coronavirus Response*, N.Y. TIMES, https://www.nytimes.com/2020/03/10/us/coronavirus-testing-delays.html (last updated Mar. 16, 2021) [https://perma.cc/5YJ9-AMGA].

240. *Id.* (reporting that as of March 13, 2020, the CDC had conducted only 8,500 tests total, while South Korea had, by late February, been able to test 10,000 people per day).

241. *See* Ryan et al., *supra* note 80.

242. *Id.*

243. On the well-established value of virus testing, see *id.*

Taking a step back, the Crimson Contagion design may have made it less likely that key decisionmakers would act on any findings from the exercise because the number of deaths was an input—fixed in advance to underscore the gravity of the situation.[244] A better-designed simulation would instead produce final reports with stark estimates for the additional number of deaths due to poor virus testing, vaccination, and other pandemic infrastructure.[245] The final reports would also ideally have produced estimates of the potential economic savings, including higher employment rates, if the government had performed better.

The point here is not to critique the COVID-19 pandemic response. We will never know what could have prevented the lost month of virus testing that enabled the original exponential growth of cases, or the subsequent shortages in testing.[246] Nonetheless, comprehensive public stress test reports would provide a paper trail that could motivate key decisionmakers to ensure that they were not responsible for ignoring such dire warnings.[247] The current limited adoption of stress tests provides leaders with too much political cover because they can always claim they were unaware of an exercise's findings or that the consequences were unclear.

If it had been faced with alarming Crimson Contagion reports released to the public, Congress may not have taken what seems in hindsight to have been a bad move: significantly cutting the CDC budget in 2019.[248] More comprehensive simulation outputs would at least have forced agency leaders to directly grasp the consequences of remaining stagnant, which could have influenced allocation of existing

244. HHS PANDEMIC SIMULATION DRAFT REP., *supra* note 68, at 9.

245. These estimates are feasible. Indeed, they are in line with the types of estimates that the government currently undertakes for varying levels of intervention, such as a faster vaccine rollout. *COVID-19 Projections*, INST. FOR HEALTH METRICS & EVALUATION, https://covid19.healthdata.org/united-states-of-america?view=total-deaths&tab=trend (last visited Feb. 17, 2022) [https://perma.cc/STD3-C8JX] (providing projections by day under various scenarios). The figures should include not only deaths caused by the simulated virus itself, but also deaths from other life-threatening issues in patients (such as those with lung disease who needed ventilators) who could not get access to critical care because hospitals were overwhelmed with the pandemic.

246. Michael D. Shear, Abby Goodnough, Sheila Kaplan, Sheri Fink, Katie Thomas & Noah Weiland, *The Lost Month: How a Failure to Test Blinded the U.S. to Covid-19*, N.Y. TIMES, https://www.nytimes.com/2020/03/28/us/testing-coronavirus-pandemic.html (last updated Oct. 1, 2021) [https://perma.cc/C5G4-YMSA] (quoting former heads of the CDC and FDA, among other officials).

247. A risk here is that this pressure causes leaders to be overly responsive. Though, given the psychological tendencies for underpreparing, more pressure has the potential to move the preparation level in the right direction.

248. Sanger et al., *supra* note 67 (mentioning funding levels and the President's budget cut proposal of nine percent in 2020).

resources. The broader point is that government leaders should make public decisions with a full picture of the risks of not preparing—both in terms of their own reputations and the well-being of society.

Well-designed public health simulations could have also prompted several straightforward, low-cost organizational improvements, any of which could have considerably lessened COVID-19's economic devastation and saved many lives, such as a more streamlined screening approval process or better organizational partnerships in place for vaccine distribution. Poorly designed public health stress tests may thus be part of the answer to an early question posed by the nation's leading pandemic epidemiologist: "Why . . . were we not able to mobilize on a broader scale?"[249]

B. Environmental Disasters

Disaster response agencies voluntarily conduct localized environmental stress test scenarios, such as an earthquake striking San Francisco or a chemical spill at a New Jersey factory.[250] Congress and state legislatures might mandate such exercises with specific timetables, public reports, inspector general oversight, and the heads of the agencies involved. These mandates would need to vary in frequency and intensity depending on the context.

In the case of the simulation that preceded Hurricane Katrina, however, the problems went beyond the delays discussed above. Local officials testified that their participation in the simulation taught them lessons that they subsequently implemented in the face of Katrina.[251] Yet the simulation prompted no such action on the part of FEMA prior to the hurricane.[252] Whereas local leaders could go back to their communities and make changes, the FEMA participants did not have that same authority because the Agency's director was not involved.[253]

In addition to the simulation's omission of the possibility that water levels would breach the city's outer walls, it also did not probe how participants would manage pre-storm evacuations.[254] The simulation's designers knew that the current evacuation preparedness

249. Shear et al., *supra* note 246 (quoting Dr. Anthony Fauci of the National Institute of Health).

250. *See* HHS PANDEMIC SIMULATION DRAFT REP., *supra* note 68, at 2–5 (an example of one such localized exercise).

251. *FEMA Simulation Comm. Hearing, supra* note 59, at 13.

252. *See id.* (investigating FEMA's simulation).

253. *Id.* at 20–21. The governor of Louisiana and other major decisionmakers did not participate, limiting simulation-related action to local entities. *Id.*

254. *Id.* at 23–24.

was insufficient to coordinate six hundred buses and convince people to leave, but that variable was eliminated due to scope management.[255]

The omission of the levee breach and evacuation variables would have been less likely with greater transparency and oversight. Instead, the out-of-sight power dynamic was between local officials who were essentially beneficiaries of the more powerful federal officials who had granted local officials' requests to fund a simulation. Adding academics, the press, and the public to the debate about what was important to include would have at least somewhat leveled the negotiation power dynamics and placed more pressure on FEMA to be thorough.

Also, the hurricane simulation failed to communicate the full stakes of acting in terms of the potential number of lives saved and economic wreckage avoided if officials had planned pre-storm evacuations and strengthened the city's storm defense infrastructure. At a minimum, with leaders involved and consequences communicated, stress tests would better set up organizations for implementing low-cost improvements, such as developing the missing emergency management plans.[256] More importantly, informed leaders are better situated to push for more resource-intense steps like rebuilding physical infrastructure.

The goal of government readiness in environmental matters is intertwined with private-sector readiness. In some areas, such as oil spills, chemical leaks, and power grid failures, no amount of government stress tests will suffice if the private actors involved are insufficiently cautious. Currently, the regime for environmental stress testing of companies varies greatly. Unlike banks and nuclear facilities,[257] electrical utilities, oil producers, and other energy companies face fewer, if any, stress test requirements.

To elaborate on the case of oil companies, the Oil Pollution Act of 1990 compels oil companies to develop plans to respond to crises.[258]

---

255. *Id.*

256. *See supra* note 102 and accompanying text (describing FEMA's lack of emergency plans).

257. PAC. GAS & ELEC. CO., *supra* note 59, at 35.

258. Oil Pollution Act of 1990, Pub. L. No. 101-380, 104 Stat. 484 (codified at 33 U.S.C. §§ 2701-2762) (allocating funds for the National Response System); *see also* 33 U.S.C. § 1321(j) (defining the requirements of the National Response System, such as requiring owners and operators of vessels to "to prepare and submit to the President a plan for responding, to the maximum extent practicable, to a worst case discharge, and to a substantial threat of such a discharge, of oil or a hazardous substance"); Hope M. Babcock, *A Risky Business: Generation of Nuclear Power and Deepwater Drilling for Offshore Oil and Gas*, 37 COLUM. J. ENV'T L. 63 (2012) (discussing how oil companies prepare to deal with the disasters that are inherent to their existence). The EPA has the statutory authority to require companies to develop risk management plans for natural disasters like hurricanes, floods, and wildfires, which can cause industrial facilities to release airborne hazardous chemicals. Emergency Planning and Community Right-to-Know Act of 1986, Pub. L. No. 99-499, 100 Stat. 1728 (codified at 42 U.S.C. §§ 11001-11050); Clean Air Act Risk Management Program, 42 U.S.C. § 7412(r)(7)(B)(ii) (requiring owners and operators of stationary sources of regulated substances beyond a threshold quantity to "prepare and implement a risk

For instance, as part of its approval permit process for the Deepwater Horizon platform, BP Oil submitted a "worst-case scenario" spill that the company forecasted as 250,000 barrels.[259] When a wellhead at BP Oil's Deepwater Horizon offshore platform severed in 2010, it discharged 4.5 million barrels of oil into the Gulf of Mexico, decimating the environment and costing local communities an estimated fifty billion dollars.[260] Yet the federal agency tasked with reviewing BP's plan failed to recognize the submitted number as a low-risk scenario.[261] Nor was BP subjected to scrutiny in terms of its potential responses to such a spill. Indeed, these permit application requirements are not even stress tests—rather, they are a scenario analysis without an accompanying simulation and evaluation.

In areas such as oil drilling, the government could occasionally require stress tests as part of obtaining a permit, once the agency is likely to grant the permit. Pre-license testing is routine in other areas of activity implicating safety, such as obtaining a license to operate a motor vehicle or become an electrician.[262] Before authorities allow a Keystone Pipeline extension or drilling off the coast of Alaska, for instance, the responsible company might be required to undertake a simulation of how they would respond to spills. The final report would include evaluative metrics such as the likelihood of containing the spill, potential damage to the ecosystem, and the economic toll to local communities.

The governmental licensing entity could then consider the comprehensive outputs of those stress tests in weighing whether to grant the application. That consideration could help improve privatized risk management by providing greater incentives to internalize the risks associated with business operations. Another way of conceptualizing this is as using stress tests to better calibrate decisions about which companies should receive risk management delegation.

---

management plan to detect and prevent or minimize accidental releases of such substances from the stationary source"); *see* Bratspies & Lamdan, *supra* note 187, at 579 (explaining how in 2017, due to Hurricane Harvey, over "forty industrial sites, including refineries and plastics plants, released dangerous amounts of hazardous pollutants into the air" in Texas and Louisiana).

259. NAT'L COMM'N ON THE BP DEEPWATER HORIZON OIL SPILL & OFFSHORE DRILLING, DEEP WATER: THE GULF OIL DISASTER AND THE FUTURE OF OFFSHORE DRILLING 84 (2011), https://www.govinfo.gov/content/pkg/GPO-OILCOMMISSION/pdf/GPO-OILCOMMISSION.pdf [https://perma.cc/4C9M-HGUV].

260. David M. Uhlmann, *After the Spill Is Gone: The Gulf of Mexico, Environmental Crime, and the Criminal Law*, 109 MICH. L. REV. 1413, 1418 (2011).

261. *See* NAT'L COMM'N ON THE BP DEEPWATER HORIZON OIL SPILL & OFFSHORE DRILLING, *supra* note 259, at 84.

262. *See generally* Eric Biber & J.B. Ruhl, *The Permit Power Revisited: The Theory and Practice of Regulatory Permits in the Administrative State*, 64 DUKE L.J. 133 (2014) (outlining the extensive use of permits in the administrative state).

Additionally, the tests could inform the conditions that the government places on licenses. Depending on the stress test results, the company might need to commit to investing in certain precautions, such as additional training, or scaling back the scope of its operations.

Of course, the additional stress tests and precautions would impose higher costs on businesses that would need to be weighed against the benefits. To limit the regulatory burden, this requirement might only make sense if the company is requesting a permit for something new, has a suspect track record, or has not recently undertaken a stress test for similar projects. But the resources devoted to the test are not necessarily economic waste and indeed may help to address a market failure of oil companies failing to internalize risks.[263] Stress tests can thus contribute to economically valuable organizational learning even for companies that routinely obtain licenses.

Stress tests could occasionally continue as the project develops, informed by any scientific or other advances in understanding risks. The regulator might provide new scenarios brought to light after the granting of the license—such as unusual potential causes of damage to the underwater oil wellhead, or an oil spill occurring during a storm. The company would then simulate its response. Other environmental areas beyond oil that could be worth stress testing include coal plant explosions that regularly release toxic fly ash into nearby communities, chemical spills by manufacturers that poison local residents, and electrical utility power blackouts that leave millions vulnerable in the midst of a snow storm.[264]

To be clear, environmental stress tests have limits in terms of the problems they can solve. Most notably, stress tests are not particularly well-suited to directly tackle arguably the world's most pressing threat: global warming.[265] Arguably, global warming progresses too gradually and causes too broad an array of harms that are difficult to link to any single agency or hazard.

Nonetheless, it would be a mistake to see stress tests as irrelevant even to this challenge. Most immediately, the kinds of disasters discussed above—oil spills, wildfires, and hurricanes—implicate climate change. For instance, natural disasters like

263. *See* Squire, *supra* note 42 (discussing environmental externalities).

264. *See, e.g.*, Patricia Helman, Student Article, *Toxic Impact: The Regulation of Coal Ash and the Influence of Big Money on Small Communities*, 23 BARRY L. REV. 41, 44–48 (2017) (discussing the dangers of fly ash to nearby communities).

265. Leaders of the world's largest businesses, governments, and nonprofits have identified global warming as one of the leading global risks. *See* WORLD ECON. FORUM, THE GLOBAL RISKS REPORT 2017, at 6 (12th ed. 2017), http://www3.weforum.org/docs/GRR17_Report_web.pdf [https://perma.cc/BCD9-6NM2].

hurricanes, floods, and wildfires can cause industrial facilities to release airborne hazardous chemicals.[266] Approximately fifteen thousand U.S. facilities must provide risk management plans to the EPA for such threats.[267] To the extent that stress tests prepare the government to prevent or contain various environmental disasters, the results would contribute to the necessarily fragmented and multifaceted effort to curb climate change.

More importantly, stress tests can encourage a more comprehensive consideration of the costs of environmental threats. As illustrated by Deepwater Horizon, some regulators and businesses irrationally (or strategically) ignore the full societal risks of development projects. If stress tests clarify and accentuate those full risks, they help address the availability heuristic and thereby encourage a more accurate decision about whether a given project is worthwhile. To the extent that such accuracy discourages some projects that would have otherwise occurred, such as building coal plants or extracting oil, stress tests could cancel projects that would have otherwise led to greater global warming. Stress tests might thus be part of the broader reckoning occurring in the private sector, which is in the process of pricing global warming into its business models.[268]

Finally, stress tests outside of environmental agencies might incorporate climate change forecasts. Federal agencies, led by a steering committee with representatives from the National Aeronautics and Space Administration, the Department of Energy, and the Environmental Protection Agency, must publish a National Climate Assessment report every four years.[269] These reports consider future-oriented scenarios and include accountability outputs. The latest report provided a worst-case scenario that temperatures could rise by as high as nine degrees by 2100, with the "risks of inaction" being hundreds of billions of dollars in additional costs and thousands more lives lost annually in the United States.[270]

---

266. Bratspies & Lamdan, *supra* note 187, at 579 (noting the dangerous release of hazardous pollutants after Hurricane Harvey in 2017).

267. *Id.* at 586 (noting that as of 2008, nine years after the final program went into place, fourteen thousand of these facilities had submitted their plans).

268. *See* Madison Condon, *Externalities and the Common Owner*, 95 WASH. L. REV. 1, 48–56 (2020) (discussing how investor pressure can help reduce greenhouse gas emissions).

269. Global Change Research Act of 1990, Pub. L. No. 101-606, 104 Stat. 3096 (codified at 15 U.S.C. §§ 2921-2961) (establishing the United States Global Change Research Program, and requiring the periodic production of a climate assessment to be presented to the president and the Congress).

270. 2 U.S. GLOB. CHANGE RSCH. PROGRAM, FOURTH NATIONAL CLIMATE ASSESSMENT: IMPACTS, RISKS, AND ADAPTATION IN THE UNITED STATES 42, 551–52 (rev. Mar. 2021), https://nca2018.globalchange.gov/downloads/NCA4_2018_FullReport.pdf [https://perma.cc//AJ4B-

These reports do not push any organization to simulate its scenarios, and thus are not stress tests. Nonetheless, those estimates are comprehensive enough to touch myriad areas of government. Since poor air quality significantly increased the number of people that died from COVID-19,[271] how should HHS adjust its simulations of pandemic-induced pneumonia deaths to account for climate change predictions? If large numbers of people migrate from their current coastal dwellings, where will they go, and at what costs to mortgage markets, insurance companies, border skirmishes, and the economy? How should the Federal Reserve respond to the financial instability implications of global warming inducing more wildfires, floods, hurricanes, and health crises?[272] The most straightforward benefit of implementing such climate change considerations into other types of stress tests would be to improve the accuracy of the underlying estimates.

That input could have the further benefit of raising greater awareness of climate change among key decisionmakers in government. With Federal Reserve officials more aware of the financial stability implications of climate change, they may be more willing to advocate for environmental action—a process that appears to have already begun.[273] Moreover, financial regulatory officials may have something to offer climate change forecasts. The National Climate Assessment omits consideration of financial stability, for instance.[274] Thus, policymakers may currently underestimate the full risks of climate change, in part due to its disconnect from financial forecasting.

In sum, the clearest and most immediately promising application of environmental stress tests is applying them more comprehensively to the prevention of natural and industrial disasters. That alone could save lives, prevent economic harms, and lessen environmental damage. A more ambitious interagency vision for

YQEK] ("Without significant greenhouse gas mitigation, the increase in global annual average temperature could reach 9°F or more by the end of this century.").

271. *See* X. Wu, R. C. Nethery, M. B. Sabath, D. Braun & F. Dominici, *Air Pollution and COVID-19 Mortality in the United States: Strengths and Limitations of an Ecological Regression Analysis*, 6 SCI. ADVANCES, Nov. 4, 2020 (finding a positive association between air pollution exposure and COVID-19 mortality rates).

272. *See* Christina Parajon Skinner, *Central Banks and Climate Change*, 74 VAND. L. REV. 1301, 1308–09 (2021) (analyzing the Federal Reserve's "legal authority to address climate change . . . through its powers in regard to monetary policy; to regulate and supervise banks; or to research and convene").

273. *See* Mary C. Daly, President & Chief Exec. Officer, Fed. Rsrv. Bank of S.F., Remarks at the Economics of Climate Change Conference: Why Climate Change Matters to Us 4 (Nov. 8, 2019), https://www.frbsf.org/our-district/files/Speech-Daly-Economics-of-Climate-Change-Conference.pdf [https://perma.cc/95N4-79TZ] (stating that climate change "can destroy wealth, exacerbate existing income inequalities, and . . . displace people permanently").

274. *See* 2 U.S. GLOB. CHANGE RSCH. PROGRAM, *supra* note 270.

applying stress tests would integrate climate change forecasts into other areas of government in a feedback loop. Interlinked stress tests could then enhance advocacy and problem solving not only for agencies but for efforts to mobilize against climate change.[275]

C. Financial Crises

Consider how redesigned stress tests might enhance financial stability. A straightforward step is recognizing that the magnitude of the problem may be much greater than even the Federal Reserve's current "worst-case" predictions. The actual drop in GDP due to COVID-19 was three times what the Federal Reserve assumed in the most negative scenario of its 2020 stress test.[276] Even though the stress test's worst-case scenario was nowhere near as bad as prior historical downturns, industry lobbied against the mild figures used by the Federal Reserve by arguing they were unrealistically negative.[277] Financial regulators could address this by using not only a highly unlikely scenario, but also occasionally a truly extreme scenario that would forecast a worse economic downturn than anything the country has ever seen.

Another problem in need of examination is that the designers of financial stress tests have focused more on changing numbers than changing events. They create different figures for the credit supply, GDP, or the length of the recession, but too often they do not imagine new types of triggers for the economic downturn—such as the triggers for the last two downturns: a housing-sector meltdown and a pandemic.[278]

The specific causes of the downturn are important beyond the numbers because they illuminate new complications. As demonstrated by both the mortgage failure and COVID-19 aftermaths, particular

275. This vision connects to democratic experimentalism. *See* Michael C. Dorf & Charles F. Sabel, *A Constitution of Democratic Experimentalism*, 98 COLUM. L. REV. 267, 345 (1998) ("The agencies are [ ] the continuing organized link between the national and the local, helping to create through national action the local conditions for experimentation, and changing national arrangements accordingly.").

276. *See* Judge, *supra* note 9 (summarizing and criticizing the Federal Reserve's approach to stress tests during COVID-19). Of course, the recovery also seems to have been faster than that modeled. For a broader discussion of stress tests as too soft, see Baradaran, *supra* note 9, at 1297. Some see the "unlikely" scenarios as wasteful. *See, e.g.*, Garrett J. Moore, *Pass or Fail? Grading the Effectiveness of Stress Tests a Decade After the Financial Crisis*, 23 N.C. BANKING INST. 333, 350 (2019) ("[D]e-regulators argue that the stress tests are too stringent because the hypothetical stressful scenarios that the Federal Reserve puts the banks through are essentially doomsday type situations that are very unlikely to happen in real life.").

277. Moore, *supra* note 276, at 350.

278. For a comprehensive summary of bank stress tests, their history, and shortcomings, as well as other helpful insights into how to improve them, see Turk, *supra* note 9, at 1719.

events can cause financial instruments and institutions to act in ways that have never happened in prior downturns with different causes. The 2020 recession brought the unprecedented combination of most nations' economies suddenly and significantly contracting, a major slump in international trade volume, large increases in debt, bonds and stocks losing value in unison (contrary to their typical inverse relationship), and treasuries becoming illiquid.[279] To create more imaginative stress tests, including unfamiliar quantitative scenarios, the tests should probe what would happen to financial institutions under more detailed real-world shocks.

To be clear, the Federal Reserve will almost surely incorporate pandemic economic indicators into its future stress tests. But in some sense that illustrates the problem, not the solution. Financial regulators must get ahead of the next crisis rather than hewing too closely to past events. The needed approach starts with learning from history but also requires imagining new threats that might emerge. Examples of exploring novel scenarios can be found in other countries. After record-high temperatures and severe droughts fueled months of devastating fires, Australia's stress test supervisor in 2020 expanded its financial institution stress tests to include climate change risks,[280] and the Netherlands has taken similar steps.[281] U.K. authorities have instructed financial institutions to take the threats created by financial technology disruption more seriously in mandatory stress tests.[282]

Scholars have also produced scenarios that regulators ignored. In 2002, Kathleen Engel and Pat McCoy wrote of a mortgage crisis

279. Irwin, *supra* note 114. The point is not that these occurrences are unprecedented, but more that the specifics of the trigger can help to push stress tests toward modeling a greater diversity of unfamiliar market dynamics.

280. *Stress Testing Assessment: Findings and Feedback*, AUSTL. PRUDENTIAL REGUL. AUTH. 3 (Feb. 21, 2020), https://www.apra.gov.au/stress-testing-assessment-findings-and-feedback [https://perma.cc/R4XN-Q5GC].

281. *See* ROBERT VERMEULEN, EDO SCHETS, MELANIE LOHUIS, BARBARA KÖLBL, DAVID-JAN JANSEN & WILLEM HEERINGA, DE NEDERLANDSCHE BANK N.V., AN ENERGY TRANSITION RISK STRESS TEST FOR THE FINANCIAL SYSTEM OF THE NETHERLANDS, 16-7 OCCASIONAL STUD. 7 (2018), https://www.dnb.nl/media/pdnpdalc/201810_nr-_7_-2018-_an_energy_transition_risk_stress_test_for_the_financial_system_of_the_netherlands.pdf [https://perma.cc/JS37-XPEX] (incorporating energy and climate change considerations into financial stress tests).

282. Jonathan Lawrence, *Bank of England Stress Tests Reveal FinTech Competition*, K&L GATES: FINTECH L. BLOG (Dec. 4, 2017), https://www.fintechlawblog.com/2017/12/bank-of-england-stress-tests-reveal-fintech-competition [https://perma.cc/Q6JS-2TU4] (describing the consideration of cybersecurity and other technological threats).

several years before the problems they identified helped spark the 2008 recession.[283] Yet regulators explicitly dismissed such warnings.[284]

One newer threat that academics have argued is worth exploring, but regulators have yet to incorporate into their stress tests, is a future when most people rely on sophisticated digital assistants for almost all spending and financial decisions.[285] If tens of millions of Americans are advised by the same Google Assistant or Apple's Siri to move their deposits from Citibank to Goldman Sachs because they can earn higher interest rates, a modern version of a bank run could result.[286]

That threat is invisible to financial regulators in large part due to the nature of past financial crises. In the 1800s and early 1900s, a major threat to financial instability was that depositors would panic and rush to withdraw their money.[287] Those bank runs can cause even healthy banks to fail because banks keep only a fraction of deposits on hand at any given time and instead use most deposits for other purposes, such as extending home loans. Lawmakers responded with Federal Deposit Insurance, which currently reimburses up to $250,000 of depositors' funds if a bank collapses.[288] That insurance promotes widespread trust in banks, thus lessening the chances that consumers will "run" on the bank out of panic.[289]

In contrast, a digital bank run would be motivated by opportunity rather than panic and is thus impervious to the regulatory architecture's main line of defense against traditional bank runs. Agencies may instead need new regulatory tools, like a circuit breaker or pause button built into digital assistants' financial advice, similar to what stock exchanges currently have when trading becomes too

283. *See* Engel & McCoy, *supra* note 133, at 1257–59.

284. *See* Kara Scannell & Sudeep Reddy, *Greenspan Admits Errors to Hostile House Panel*, WALL ST. J. (Oct. 24, 2008, 12:01 AM), https://www.wsj.com/articles/SB122476545437862295 [https://perma.cc/RF9Z-7Q8X].

285. *See, e.g.*, Tom Baker & Benedict Dellaert, *Regulating Robo Advice Across the Financial Services Industry*, 103 IOWA L. REV. 713, 714–15 (2018); Rory Van Loo, *Making Innovation More Competitive*, 65 UCLA L. REV. 232, 234 (2018).

286. Rory Van Loo, *Digital Market Perfection*, 117 MICH. L. REV. 815, 868 (2019).

287. BARR ET AL., *supra* note 111, at 42–48.

288. Anthony J. Casey & Eric A. Posner, *A Framework for Bailout Regulation*, 91 NOTRE DAME L. REV. 479, 480 (2015).

289. As another example of a technological stress test, some have proposed probing how the financial system would withstand a cyberattack. *See, e.g.*, Tom C.W. Lin, *Financial Weapons of War*, 100 MINN. L. REV. 1377, 1431 (2016) ("Policymakers should design advanced technological stress tests to assess the information technology infrastructure of systemically important private and public financial institutions and agencies."); *see also, e.g.*, Nizan Geslevich Packin, *Too-Big-to-Fail 2.0? Digital Service Providers as Cyber-Social Systems*, 93 IND. L.J. 1211 (2018) (arguing for similar attention to critical technological service providers as is paid to big banks for instability).

volatile.[290] These kinds of futuristic disasters that the current legal architecture is inadequate to prevent are the types that should be modeled as part of imaginative stress tests.

Outlandish scenarios rooted in real-world trends are worth considering even though—like many past triggers of financial crises—they are individually unlikely to occur. The nature of new triggers for financial downturns is that they are implausible until they actually happen. For instance, nobody saw Reddit, Robinhood, and other technologies driving the kind of wild fluctuations of GameStop and other stocks in early 2021.[291] Stress tests should consider unusual scenarios driven by the diverse digital platforms influencing finance, and it is in some ways shocking that—decades after the internet became mainstream—they still do not.

That approach not only would prepare for as many specific scenarios identified as possible but may also indicate actions that would be relevant to a broader set of unforeseen triggers. For instance, if bank regulators could pause automated digital assistants in the case of emergency, it would also guard against a scenario in which hackers were suddenly using those digital assistants to withdraw funds from millions of accounts—even if that specific scenario was not the one that had led lawmakers to extend the pause authority. Imaginative stress tests thus increase the chances that government is prepared even for threats different from those in the unlikely scenarios.

So far, the discussion has focused on making existing bank stress tests more imaginative.[292] However, existing stress tests focus on the bank's responses. Another key shift is implementing stress tests of the Federal Reserve and other key financial regulators, rather than just of banks. The idea would be to bring to the financial system what currently exists in the military, where the Pentagon simulates what command headquarters would do in the face of a nuclear attack.[293]

---

290. Van Loo, *supra* note 286, at 879–80 (discussing the NYSE's pause feature and potential use of slow-down mechanisms by regulators).

291. *See* Matt Phillips & Taylor Lorenz, *'Dumb Money' Is on GameStop, and It's Beating Wall Street at Its Own Game*, N.Y. TIMES, https://www.nytimes.com/2021/01/27/business/gamestop-wall-street-bets.html (last updated Feb. 25, 2021) [https://perma.cc/QSZ2-U2BQ] (discussing how social media facilitated massive increases in GameStop share prices).

292. Stress tests of banks indirectly help agencies assess their performance in preventing a crisis, because if a bank fails, then the regulator has failed in one of its primary missions.

293. For an excellent proposal for improving financial regulators' training for crises through simulations, see Crawford, *supra* note 9, at 161–74. For valuable ideas on how to reform bank stress tests with more of a public component, see, for example, Baradaran, *supra* note 9, at 1319, which argues that bank stress tests should draw from military war games exercises; Turk, *supra* note 9, at 1701, which proposes that the stress test rules written by financial regulators be stress tested; and Weber, *supra* note 9, at 2242, which offers a large array of improvements to financial stress tests.

To accomplish that goal, the law could mandate that authorities create well-developed plans with specific actions that agency leadership and other employees would take in a crisis. Which meetings would be called with which external stakeholders? When every second counts and multiple decisions must be made, how should the Federal Reserve Chair prioritize issues? It was alarming that financial regulators—whose primary goal is financial stability—did not have tested emergency plans in place in the last financial crisis.[294]

In a financial crisis, like in a pandemic, time is of the essence.[295] Stocks may plunge precipitously, mass layoffs may occur, and healthy companies may fold before Congress can act.[296] Agency and congressional staff will not have time, once an emergency arrives, to thoroughly research best practices from all past calamities, domestic and abroad, and craft responses. An exercise addressing those challenges would therefore require financial regulators to undertake a fire drill by forcing them to hold meetings and make decisions as if a given digital bank run, deadly pandemic, or unprecedented economic downturn had occurred. The agency's futurist or innovation officer would provide scenarios for stress tests. If mandated by Congress, the inspector general or some outside entity could produce a report card for the agency's performance, including how many banks failed, or trillions of dollars in household wealth were lost, due to poor judgment.

One of the outputs of financial stress tests could be draft legislation on whatever stimulus, modification to financial institution bankruptcy, or other intervention might be needed in light of the particular scenario tested. Draft legislation is crucial in financial regulation because lawmakers pressed for time are more likely to rely on bills drafted by lobbyists.[297] As the economic crisis of 2020 arrived, Congress had just days to write the biggest stimulus package the country has ever adopted, leaving the $2.2 trillion legislation especially susceptible to influence by the lobbyists that bombarded the process.[298] The 2008 bailout package for the auto industry and financial

294. *See supra* Section I.B.3 (discussing limitations of financial regulatory stress tests).

295. Andrew P. Atkins, *The AIG Bailout: Constraining the Fed's Discretion*, 14 N.C. BANKING INST. 335, 358 (2010) (emphasizing "speed and flexibility" as crucial in financial emergencies).

296. *Id.*

297. Alexander Hertel-Fernandez, *Who Passes Business's "Model Bills"? Policy Capacity and Corporate Influence in U.S. State Politics*, 12 PERSPS. ON POL. 582, 587 (2014).

298. *See* Erica Werner, Mike DeBonis & Paul Kane, *Senate Approves 2.2 Trillion Coronavirus Bill Aimed at Slowing Economic Free Fall*, WASH. POST (Mar. 25, 2020), https://www.washingtonpost.com/business/2020/03/25/trump-senate-coronavirus-economic-stimulus-2-trillion/ [https://perma.cc/44EH-T74Q] ("Lawmakers and the White House were bombarded with lobbyists and special-interest groups seeking assistance during the negotiations . . . .").

institutions similarly unfolded in days, prompting considerable criticism once it became clear how much of the package had benefitted large banks and their senior executives.[299]

Thus, although not essential, the participation of congressional staffers in some financial regulatory stress tests is worth considering, particularly those working for legislators on the U.S. House Committee on Financial Services and the U.S. Senate Committee on Banking, Housing, and Urban Affairs.[300] Off-the-shelf draft statutes and regulatory plans—developed more imaginatively without time pressure—would provide stronger foundations for rapid adjustments when the next financial collapse arrives. Financial crises thus illuminate a more expansive vision for stress tests' broader potential as a public governance tool beyond administrative agencies.

D. Stress Testing Tech and Beyond

This Article has identified many contexts in which laws and administrative agencies mandate stress tests of private entities. Those mandates reach banks, nuclear power plants, and hospitals, among others. The discussion has also briefly explored whether other areas, such as environmental and trade regulation, merit more robust stress testing of private entities. A question surfaced by these discussions is what other private-sector contexts might be candidates for stress tests.

In identifying new contexts, two features stand out as indicating that mandated stress tests have potential to help: risks of great societal harms combined with externalities that make businesses unlikely to adequately internalize those risks. Another factor that could be persuasive to some, but is difficult to determine, is the degree to which the entity plays a quasi-public role. Some tech companies, such as Google and Facebook, may have become so central to society, for everything from information access to elections, that it is worth thinking of how stress tests might be integrated into their governance.[301]

---

299. *See* Adam J. Levitin, *In Defense of Bailouts*, 99 GEO. L.J. 435, 438 (2011) (commenting that financial firms receiving government bailouts, particularly AIG, were criticized for then providing lavish executive bonuses).

300. *See* Martha L. Cochran, David F. Freeman & Helen Mayer Clark, *Money Market Fund Reform: SEC Rulemaking in the FSOC Era*, 2015 COLUM. BUS. L. REV. 861, 906 (noting that as a condition of Dodd-Frank the Federal Reserve Board of Governors must report to these two committees on any emergency lending program within seven days of initiation).

301. On the broader point of how risk management may be necessary to address the growing risks posed by some tech companies and the information age more broadly, see, for example, Packin, *supra* note 289, at 1215 (calling for new risk management of digital service providers such as Google, Amazon, Apple, and Facebook); and Cohen, *supra* note 17, at 391 ("If regulators are to

One subject matter area that may fit these basic criteria is cybersecurity.[302] In 2021, hackers shut down one of the largest U.S. pipelines until the owner, Colonial, paid a ransom.[303] Since the pipeline provides forty-five percent of the East Coast's fuel, the event disrupted transportation.[304] This attack also showed how hackers might disable critical infrastructure upon which the nation depends. Criminal groups have launched similar ransomware attacks against hospitals and manufacturers, but private entities typically try to cover up those events, which may cause people to underestimate the extent of the threat.

One challenge in drawing any definitive conclusions about the need for stress tests for cybersecurity is that businesses surely have incentives to guard against such attacks. Many technology companies have voluntarily adopted stress tests to probe their defenses against hackers.[305] Moreover, it would be difficult to determine when that incentive is adequate, leading to the risk of overprescribing stress tests.

However, companies also arguably have incentives to avoid legal violations, to escape monetary penalties, and to minimize reputational harm. Yet strong incentives have not always proved sufficient to deter misbehavior. Additionally, there may be economies of scale in a governmental stress testing program, which would spread the costs of designing imaginative scenarios and consequential outputs across many different private-sector organizations.

Another business context that officials have recently analyzed for stress tests is supply-chain management. The COVID-19 pandemic illustrated the vulnerability of supply chains in pharmaceuticals, high-capacity batteries, semiconductor chips, and critical minerals.[306] Whereas supply-chain breakdowns decimated most auto manufacturers, Toyota withstood shortages during the pandemic

develop a more effective set of tools for avoiding systemic breakdown, more comprehensive engagement with threat modeling and its pitfalls is essential.").

302. Cybersecurity has entered into the academic conversations about bank stress tests. *See, e.g.*, Richard J. Herring & Til Schuermann, *Objectives and Challenges for Stress Testing*, *in* HANDBOOK OF FINANCIAL STRESS TESTING (J. Doyne Farmer, Alissa Kleinnijenhuis, Til Schuermann & Thom Wetzer eds., forthcoming 2022) (proposing cybersecurity attention in stress tests).

303. David E. Sanger, Clifford Krauss & Nicole Perlroth, *Cyberattack Forces a Shutdown of a Top U.S. Pipeline*, N.Y. TIMES, https://www.nytimes.com/2021/05/08/us/politics/cyberattack-colonial-pipeline.html (last updated May 13, 2021) [https://perma.cc/LP42-N66Z].

304. *Id.*

305. Nathan Alexander Sales, *Privatizing Cybersecurity*, 65 UCLA L. REV. 620, 634, 686 (2018) (describing "bug bounty programs, in which software vendors and other companies offer modest cash rewards to researchers who report vulnerabilities in their products and then issue patches for the flaws" and how bug bounties can be paid to outside hackers or in-house employees).

306. Tausche, *supra* note 8.

because it had already failed a real-life disaster in the form of a devastating 2011 earthquake that knocked its semiconductor chip supplier offline.[307] It later decided to stockpile chips and other supplies in preparation for unexpected disruptions.[308] White House officials have considered pushing supply-chain stress tests to encourage similar preparedness in other companies.[309]

It is also worth noting that stress tests of businesses need not always be mandated by law or government officials. The psychological problem of leaders paying insufficient attention to rare threats is something that shareholders or executives may also want to address. The board or the CEO can require stress tests of the firm if they are concerned that managers are insufficiently prepared for rare events, even when the business does not build airplanes or nuclear reactors. One area for future research is thus how stress tests might improve corporate governance.

CONCLUSION

Stress tests of agencies and firms have become a widespread tool in the administrative state. Agencies use them to lessen the chances of devastation from pandemics, financial meltdowns, environmental tragedies, nuclear attacks, and many other dangers. In theory, these simulations can further the core administrative law goal of accountability by addressing well-known psychological shortcomings that cause decisionmakers to inadequately identify and weigh risks. However, too often the tests are secretive, backward-looking, or inconsequential. Agency leaders can exercise their discretion to voluntarily improve those shortcomings, but there is little reason to expect that to happen systematically.

Two years after the 2008 financial crisis, Congress passed the most sweeping financial reform legislation in generations, establishing what Federal Reserve Chair Jerome Powell described as the "cornerstone" of financial regulation: annual stress testing of systemically important financial institutions.[310] The COVID-19 pandemic and accompanying recession may present a similar rare opportunity for a large-scale overhaul. If so, as part of that legislation

307. River Davis, *How Toyota Steered Clear of the Chip Shortage Mess*, BLOOMBERG BUSINESSWEEK (Apr. 7, 2021, 4:00 PM), https://www.bloomberg.com/news/articles/2021-04-07/how-toyota-s-supply-chain-helped-it-weather-the-chip-shortage [https://perma.cc/G4AZ-AZ6J].

308. *Id.*

309. Tausche, *supra* note 8.

310. Powell, *supra* note 58; Dodd-Frank Wall Street Reform and Consumer Protection Act, Pub. L. No. 111-203, § 165(i), 124 Stat. 1376, 1430 (2010) (codified as amended at 12 U.S.C. § 5365).

Congress should consider mandating transparent, imaginative, and consequential stress tests of public health agencies. Moreover, lawmakers may want to seize the national sense of vulnerability to go further and improve upon the widespread use of stress tests across the administrative state.

As crucial as those specific reforms are to position the government for success in its most important job of keeping Americans safe, this Article's thesis may have even larger implications. The administrative state has been called the fourth branch of government.[311] One of the core functions of each branch is to provide a check on the others. Well-designed stress tests can signal governmental failure in the existing system of laws, delegated authority, and executive administration. Stress tests thus can provide not only accountability for agencies but also a means of signaling when the law as written by the legislature or implemented by the executive fails to safeguard society from its biggest threats.

Greater attention is thus needed to stress tests and related tools that might spur an inert state to action. More gradual threats, such as climate change and the social media spread of disinformation, may benefit from simulations of future scenarios that include elements of stress tests. It is worth exploring, for instance, whether bureaucrats and even lawmakers should participate in simulations that would force them to confront a future planet nine degrees hotter and with significantly more toxic air. Since history has taught that government leaders often cannot be trusted to take necessary steps absent a looming threat, the law should sometimes push them to act in the calm between crises.

311. Peter L. Strauss, *The Place of Agencies in Government: Separation of Powers and the Fourth Branch*, 84 COLUM. L. REV. 573, 578 (1984) ("Almost fifty years of experience has accustomed lawyers and judges to accepting the independent regulatory commissions, in the metaphor, as a 'headless "fourth branch" ' of government.").