
Article

A Theory for Deliberation-Oriented Stress Testing Regulation

Robert Weber[†]

Introduction	2237
I. Origins of Stress Testing as the Analysis of Failure	2244
A. Structural and Mechanical Engineering.....	2245
B. Systems Reliability Analysis	2248
II. How Do Stress Tests Work in Finance?	2249
A. Why Stress Is a Promising Regulatory Tool.....	2249
B. Stress Tests Are a Key Component of Existing Risk Management Systems.....	2250
1. Brief Introduction to Risk Management Systems	2251
2. Brief Introduction to Value-at-Risk, a Building Block of Risk Management Systems	2253
C. A Taxonomy of Financial Stress Testing Techniques	2258
1. Firm-Conducted Stress Tests	2261
a. Sensitivity Analysis	2261
b. Portfolio-Driven Scenario Analysis.....	2261
c. Event-Driven Scenario Analysis	2262
d. War Games	2263
2. Regulator-Conducted Stress Tests.....	2266
III. Elaborating the Distinction between Deliberation- Oriented Stress Testing and Assurance-Oriented	

[†] Wellspring Professor of Entrepreneurship, Economic Development, and Business Law, University of Tulsa College of Law. I thank the attendees and participants at the National Business Law Scholars Conference at Ohio State University Moritz College of Law, the 2013 University of Connecticut School of Law Junior Scholars Workshop on Financial Services Law, and the “Financial Crisis, Financial Institutions and Governance: Causes, Consequences and Reforms” panel at the 2013 Annual Law & Society Meeting. Special thanks go to Mehrsa Baradaran, John Crawford, Cristie Ford, Anna Gelpert, Kristin Johnson, Patricia McCoy, David Min, and Daniel Schwarcz for helpful comments. Copyright © 2014 by Robert Weber.

2014]	<i>STRESS TESTING</i>	2237
Stress Testing.....		2268
A. Assurance-Oriented Stress Testing.....		2270
B. Deliberation-Oriented Stress Testing		2273
C. Four Hypothetical Case Studies of Stress Testing		2276
1. “Bank of Hazard”		2276
2. “Bank of Prudence”		2277
3. Isolated Stress Test of GMDB Exposures		2278
4. Ongoing Program of Dialogic Stress Testing of GMDB Exposures.....		2279
IV. A Brief History of the Regulation of Financial Stress Testing		2280
A. Two Improbable Early Stress Testing Initiatives		2280
B. The Development of Stress Testing Regulation from 1990-2006		2284
C. A New Phase of Post-2008 Stress Testing Regulation		2290
1. Regulator-Conducted Assurance-Oriented Stress Testing.....		2291
2. Basel Committee Principles: An Overture to Deliberation-Oriented Stress Testing?.....		2299
V. Three Principles to Guide the Use of Regulation to Institutionalize Deliberation-Oriented Stress Testing		2301
A. Management-Based Regulation: Understanding Why Stress Testing Makes Sense as a Regulatory Mechanism.....		2303
B. An Industrial Morality of Quantitative Skepticism		2308
C. High-Reliability Organizations: Regulators Already Have Prototypes and Need Not Create Anything New		2313
Conclusion.....		2324

INTRODUCTION

Of the many recent changes in the legal-regulatory regimes affecting financial institutions, the advent of stress testing as a key regulatory tool deserves special attention on account of its novelty and unique potential.¹ To most Americans—and indeed,

1. See Chester S. Spatt, *Regulatory Conflict: Market Integrity vs. Financial Stability*, 71 U. PITT. L. REV. 625, 627 (2010) (“One of the most interesting innovations to emerge in the bank supervision model during the financial market crisis is the use of stress tests.”); Patrick Jenkins & Brooke Masters, *Again Under Strain*, FIN. TIMES, Jul. 8, 2011, at 7 (“[S]ince the financial crisis

to most lawyers—the mention of a stress test likely prompts thoughts of a visit to the cardiologist before it suggests anything about the financial system.² Those trained in engineering will have a different view of stress testing in mind involving the study of how physical objects such as bridges or airplane wings respond to applied forces. The multitude of disciplinary contexts in which stress testing concepts have gained currency testifies to the general usefulness of stress as a diagnostic tool. For those involved in financial markets, the appeal is intuitive: after an epochal series of stressed market events convulsed financial markets and systems at the end of the previous decade, of course it is advisable for regulators and firms to spend more time thinking about stress. However, this intuitive appeal presents a potential problem if, in their effort to incorporate stress testing into regulatory regimes, lawmakers and regulators overzealously clump together techniques and models of stress testing without taking account of how they diverge in important ways.

In a broad and interdisciplinary sense, a stress test is any analytic exercise designed to gauge how changes in variables, usually of a dramatic or “stressed” nature, affect a test subject in ways that are relevant to the subject’s performance, and in particular its susceptibility to failure. Ultimately, the purpose of any stress test is to gain familiarity with the means by which those variables negatively impact the behavior of the test subject.³ In the financial area, stress tests can help us to understand how an institution or system would respond to severe, yet plausible, stressed market conditions such as low economic output, high unemployment, stock market crashes, liquidity

. . . stress-testing [has] become a vital part of the regulatory arsenal.”)

2. Cardiology stress tests were frequently invoked by analogy after the announcement that U.S. banking regulators would be conducting stress tests at the 19 largest U.S. bank holding companies in connection with the Treasury’s Supervisory Capital Assessment Program. See Press Release, Timothy Geithner, U.S. Treas. Sec’y, Secretary Geithner Introduces Financial Stability Plan (Feb. 10, 2009), available at <http://www.treasury.gov/press-center/press-releases/Pages/tg18.aspx> (“First, we’re going to require banking institutions to go through a carefully designed comprehensive stress test, to use the medical term.”); David Wessel, *Bank Checkup Also Tests Regulators*, WALL ST. J., Apr. 16, 2009, <http://online.wsj.com/news/articles/SB123983475012122683> (“In medicine, a stress test is when the doctor puts you on a treadmill to check your heart. . . . Today, the stress test the [U.S. regulators] are conducting at the nation’s largest banks is more than a test of the patients’ health. It is a test of the government’s ability to restore confidence . . .”).

3. See, e.g., AUGUSTO J. DURELLI, APPLIED STRESS ANALYSIS, at xii (1967).

shortages, high default rates, and failures of large counterparties.⁴ The results of stress tests shed light on the tension points and weak links in portfolios and systems that could create extraordinary but plausible losses.⁵

Stress testing in the financial industry has been a key component of financial firms' internal risk management departments for decades.⁶ What is new, however, is the zeal with which lawmakers and regulators have looked to stress testing as a regulatory technique. As lawmakers have increasingly emphasized stress testing as a tool of bank regulation, they have failed to elaborate several fundamental distinctions among stress testing techniques and programs. Only with a proper understanding of its purposes and limitations is a stress testing regulatory program likely to prove an effective policy tool.

Before presenting a normative case for how policymakers should design the law and regulation of stress testing, this Article undertakes three descriptive tasks to better understand the diverse set of techniques labeled as stress tests. First, it briefly describes the origins of stress testing practices in the engineering discipline as the study of how structures and systems experience failure. By placing stress tests in their historical and conceptual context, their purposes and promises should become clearer. Second, it presents a rough typology of financial stress test techniques and discusses the distinctions among them, highlighting in particular the extent of their reliance on hands-on management involvement and the usage of quantitative statistical models that have come to predominate in financial risk management. Financial firm risk managers conduct a varied array of exercises that fit the stress test label, including sensitivity analysis, hypothetical scenario analysis, historical simulation, "war game" direct event simulation, and "reverse" stress tests. Regulators also conduct their own stress tests of the financial system or individual institutions, either on an ad hoc or periodic basis. Third, it surveys the law and regulation of stress testing up to the present.

4. See Mathias Drehmann, *Stress Tests: Objectives, Challenges and Modeling Choices*, 2 SVERIGES RIKSBANK ECON. REV. 60, 63 (2008).

5. See PHILIPPE JORION, FINANCIAL RISK MANAGER HANDBOOK 266–68 (5th ed. 2009).

6. See, e.g., MICHEL CROUHY ET AL., RISK MANAGEMENT 232–40 (2001); JAMES T. GLEASON, RISK: THE NEW MANAGEMENT IMPERATIVE IN FINANCE 190–93 (2000); GRP. OF THIRTY, DERIVATIVES: PRACTICE AND PRINCIPLES 11–12 (1993); DAVID SHIRREFF, DEALING WITH FINANCIAL RISK 38 (2004).

The Article also makes two theoretical contributions. First, it develops an overarching distinction between two alternative institutional orientations towards stress testing that has largely escaped scrutiny: *assurance-oriented stress testing* versus *deliberation-oriented stress testing*. These disparate orientations characterize the corporate governance structures and firm attitudinal settings within which the different stress testing techniques, discussed above, are put to use. It goes on to argue that the latter in particular holds great promise to counteract some of the trenchant problems in contemporary finance. Second, it presents a three-part framework, summarized below, to guide efforts to implement a regulatory regime that encourages deliberation-oriented stress testing.

When using stress testing as an assurance tool, banks—or regulators in coordination with banks—assume a partial equilibrium model of the financial system; privilege static scenarios; draw from audit culture; rely on historical precedents in setting scenarios; examine test variables in isolation; require precise estimation; design tests with compliance and verification concerns in mind; and emphasize the communicative function of the tests. A recent, salient example of assurance-oriented stress testing was the highly publicized stress testing program conducted by the U.S. government in 2009, labeled the Supervisory Capital Assessment Program (SCAP). The SCAP exercise subjected the nineteen largest U.S. bank holding companies to stress tests designed to gauge likely losses through 2010 given a “baseline” scenario and a “more adverse” scenario.⁷ This manifestation of stress testing receives the most attention from industry and commentators, but the familiarity of these uses of stress tests disguises some of their limitations.

These limitations emerge most clearly when assurance-oriented stress testing is compared with the second, and more promising, orientation towards stress testing, which I label *deliberation-oriented stress testing*. *Deliberation-oriented stress testing* privileges dynamic scenarios; draws from business operations culture; relies on imagination; considers the interactivity of tested variables; remains open to uncertainty; and is motivated by governance concerns. In this manifestation, stress

7. See CONG. OVERSIGHT PANEL, JUNE OVERSIGHT REPORT: STRESS TESTING AND SHORING UP BANK CAPITAL 10–26 (2009); Edmund L. Andrews & Eric Dash, *Government Offers Details of Bank Stress Test*, N.Y. TIMES, Feb. 25, 2009, <http://www.nytimes.com/2009/02/26/business/economy/26banks.html>. For more on the SCAP stress tests, see *infra* Part IV.C.

testing emerges less as a mechanical assurance tool and more as an aspirational internal corporate governance norm. One recent example of a *potential* deliberation-oriented stress testing initiative is subsection (i)(2) of section 165 of the Dodd-Frank Act, which—without defining stress test or identifying the necessary parameters to stress—requires most banks subject to U.S. banking regulation to conduct stress tests on their own.⁸

The great promise of deliberation-oriented stress testing is that it will foster a more mindful decisional infrastructure within financial institutions and institutionalize the continuous deliberation on failure, catastrophe, and stress. Given the roots of stress testing in the engineer's study of *failure modes*, it is a ready conceptual fit in the context of the financial regulatory regime, the primary objectives of which are (1) *to prevent failures* in core financial utility functions such as the payment system and the monetary policy transmission; and (2) *to limit failures* of individual institutions that result in recourse to government safety nets.⁹ Because it affects the corporate governance infrastructure of financial institutions, deliberation-oriented stress testing might facilitate regulatory efforts to promote these regulatory objectives under conditions of increasing uncertainty, volatility, and authentic complexity.¹⁰

Conceptualizing stress testing regulation as a corporate governance matter makes sense. Unlike the physical systems with which engineers are concerned, financial professionals and regulators must take into account human and organizational decision making. And deliberation-oriented stress testing might counteract decisional pathologies—such as the disqualification

8. 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). The mandate to conduct internal stress testing pursuant to subsection (i)(2) applies to (a) financial institutions designated by the Financial Stability Oversight Council (FSOC) as systemically significant and (b) bank holding companies with assets in excess of \$10 billion. *See id.* §§ 113, 165(i).

9. *See* RICHARD S. CARNELL ET AL., *THE LAW OF BANKING AND FINANCIAL INSTITUTIONS* 53–61 (2009).

10. *See generally* Robert Weber, *Structural Regulation as Antidote to Complexity Capture*, 4 AM. BUS. L.J. 643 (2012) (arguing that in many respects contemporary financial systems are characterized by complexity rather than mere complicatedness and discussing the implications); *see also* CHRISTINE PARKER, *THE OPEN CORPORATION: EFFECTIVE SELF-REGULATION AND DEMOCRACY* 29 (2002) (“The experience of command and control shows that it is not reasonable, practical or efficient for external legislatures and regulators to be solely responsible for determining how organizations should manage social issues. The design and enforcement of regulation to govern every potential social dilemma facing business is simply not achievable.”).

heuristic, the outcome bias, the overconfidence bias, and the hindsight bias—that inhibit organizational learning and unavoidably affect how capital is allocated throughout the firm. For reasons discussed below, these tendencies often result in key corporate decision makers overlooking accumulating evidence of anomalies, overestimating the validity of their expectations, and struggling to distinguish narrowly-avoided catastrophes from outright successes. Psychologists have made further recent discoveries that people interpret near-miss events—that is, narrowly avoided catastrophes—in ways that actually *lower*, rather than heighten, their perceptions of risk.¹¹ This troubling finding suggests the counterintuitive result that narrowly avoiding catastrophe makes us less likely to perceive our activities as risky and might even reduce our ability to engage in learning. These decisional pathologies are prevalent in the context of financial firm governance, and are among the most severe threats to financial stability because they inhibit managers' and board members' ability to understand risk. The 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 and to normalize (rather than problematize) unexpected events that might be weak-signal harbingers of future catastrophe.

Using public regulatory power to institutionalize deliberation-oriented stress testing within financial firms presents unique challenges. The open-ended and indeterminate character of its outputs is a poor fit with existing conceptions of administrative law. This Article recommends a three-part framework to guide implementation of regulatory systems designed to promote deliberation-oriented stress testing. First, regulators should understand their task as involving *management-based regulation*—a regulatory approach that acts on corporate planning processes rather than specific corporate actions or outputs.¹² Second, they should encourage the further development of the “quantitative skepticism” that Anette Mikes, in her field research with bank risk management departments, has

11. See Robin L. Dillon & Catherine H. Tinsley, *How Near-Misses Influence Decision Making Under Risk: A Missed Opportunity for Learning*, 54 MGMT. SCI. 1425, 1436 (2008).

12. See Cary Coglianese & David Lazer, *Management-Based Regulation: Prescribing Private Management To Achieve Public Goals*, 37 LAW & SOC'Y REV. 691, 691 (2003).

identified as one of two predominant calculative cultures within large financial institutions.¹³ Deliberation-oriented stress testing will have an easier time gaining traction in a quantitatively skeptical risk management department. Third, in their interactions with regulated firms, regulators should draw from the examples provided by so-called high-reliability organizations (HROs) such as nuclear power plants, aircraft carriers, air traffic control systems, and wildfire-fighting units. Management and organizational science scholars have lauded the ability of HROs to maintain reliable, resilient performance even in conditions of uncertainty and volatility.¹⁴ Put another way, the success of stress testing as a regulatory tool may well depend on whether the regulated institutions implement HRO decision-making norms throughout the enterprise.¹⁵

The effort to build a more mindful decisional infrastructure in the corporate governance of financial firms is at bottom about building not only technical knowledge but also imaginative capabilities. The nature of this task recalls the recommendation of the so-called 9/11 Commission, charged with making suggestions on how to avoid another catastrophic terrorist attack on major domestic infrastructure, that governmental agencies “routiniz[e], even bureaucratiz[e], the exercise of imagination.”¹⁶ The corporate governance dimension of deliberation-oriented stress testing should be similarly conceptualized

13. For an extensive discussion of Mikes’s research, see *infra* notes 56–58, 377–87 and accompanying text.

14. See CONSTANCE PERIN, *SHOULDERING RISKS: THE CULTURE OF CONTROL IN THE NUCLEAR POWER INDUSTRY*, at ix (2005); JOSEPH V. REES, *HOSTAGES OF EACH OTHER: THE TRANSFORMATION OF NUCLEAR SAFETY SINCE THREE MILE ISLAND* 123–50 (1994) (discussing learning from the experience of Three Mile Island in the nuclear power sector); KARL E. WEICK & KATHLEEN M. SUTCLIFFE, *MANAGING THE UNEXPECTED: RESILIENT PERFORMANCE IN AN AGE OF UNCERTAINTY* 1 (2007); Jos A. Rijpma, *From Deadlock to Dead End: The Normal Accidents—High Reliability Debate Revisited*, 11 *J. CONTINGENCIES & CRISIS MGMT.* 37, 37 (2003).

15. Of course, society almost assuredly has a lower risk tolerance for failure of a nuclear power plant than for a large bank or insurer. The point is not that financial institutions are just like nuclear power plants; instead, the point is that in both cases failure can impose externalities that justify a regulatory intervention into the private ordering of corporate decision making.

16. NAT’L COMM’N ON TERRORIST ATTACKS UPON THE U.S., *FINAL REPORT OF THE NATIONAL COMMISSION ON TERRORIST ATTACKS UPON THE UNITED STATES* 344 (2004) [hereinafter 9-11 COMM’N REPORT], available at <http://www.9-11commission.gov/report/911Report.pdf>. Analogously, psychologist Gary Klein has identified that successful firefighters have rich imaginative lives. See GARY KLEIN, *SOURCES OF POWER: HOW PEOPLE MAKE DECISIONS* 1–39 (1998).

as an effort to cause financial firms to “expand [their] conception of the possible” in an effort to better understand the causal environments, often characterized by uncertainty and volatility, that affect the achievement of firm objectives.¹⁷

Part I of the Article briefly describes the origins of stress testing practices in the engineering discipline. By placing them in their historical and conceptual context, the purposes and promises of stress testing should become clearer. Part II describes stress testing in the financial industry, situating such techniques within a broader risk management framework of which they form a part and introducing a taxonomy of the prevalent techniques labeled as “stress tests” in finance. Part III elaborates the distinction between assurance-oriented stress testing and deliberation-oriented stress testing and argues that the latter model holds more promise for promoting financial stability. In so doing, it also presents four fictional case studies to illustrate differences between these two orientations. Part IV sets forth a brief history of the regulation of stress testing practices of financial institutions and finds that assurance-oriented stress testing has, to date, predominated the field. Part V advocates in favor of the three-part framework—resting on the pillars of management-based regulation, quantitative skepticism, and high-reliability organization principles—that will serve as a useful guide for regulators to promote deliberation-oriented stress testing.

I. ORIGINS OF STRESS TESTING AS THE ANALYSIS OF FAILURE

What is a stress test? The details of the answer to that question will depend on the discipline to which you direct the question. Nevertheless, common conceptual threads run through all applications of stress tests, the most obvious of which is a focus on the *causes of failure*.¹⁸ When conducting the test, the analyst applies forces to the test object, whether experimentally or analytically, up to or approaching some point of failure in order to determine how the forces affect the test object’s properties.¹⁹ Failure, as used in this context, should be understood as the non-fulfillment of an objective, and therefore

17. Tony Jackson, *Turning the Screw Back to 1973—Or Perhaps Further*, FIN. TIMES, Nov. 25, 2007, <http://www.ft.com/intl/cms/s/0/93d95d0e-9b7b-11dc-8aad-0000779fd2ac.html>.

18. See, e.g., DURELLI, *supra* note 3, at x–xii.

19. See, e.g., *id.*

requires consideration of the test object's purpose. Thus, a mid-flight fracture of an airplane wing would certainly constitute a failure, but so too might a gradual deformation of the wing's edge due to shear stress that does not, at present, compromise its ability to fly but which nevertheless could lead to an air disaster if left unrepaired. By better understanding the historical genesis of stress testing methodologies, we can situate the contemporary uses of stress testing in financial regulation in their conceptual context.

A. STRUCTURAL AND MECHANICAL ENGINEERING

This focus on how stress can lead to failure has deep roots in the engineering discipline. The first stress tests were performed by the historical predecessors of structural and mechanical engineers, who sought to determine the strength of the materials they used to build structures.²⁰ Nowadays, the term stress connotes any “overpowering pressure of some adverse force or influence.”²¹ When the term is used in financial regulation or legislation, this definition is intended. But for engineers, stress has a specialized meaning as a measure of the “intensity of the internally distributed forces or components of forces that resist a change in the form of a body.”²²

Analysis of how things fail contributes to our understanding of how things work. Writing of the advances in scientific knowledge during the Scientific Revolution, historian William Rosen notes that “understanding didn't progress by looking for truth; it did so by looking for mistakes.”²³ Henry Petroski, a Duke University engineering professor, applies Rosen's observation to his own discipline when he notes that “engineering understanding did not progress by looking at successes; it did so by looking at failures.”²⁴ Petroski emphasizes how the engi-

20. See ARNOLD W. HENDRY, *ELEMENTS OF EXPERIMENTAL STRESS ANALYSIS*, at vi (1964).

21. THE OXFORD ENGLISH DICTIONARY 885 (2d ed. 1989) (defining “stress”).

22. See HARMER E. DAVIS ET AL., *THE TESTING OF ENGINEERING MATERIALS* 17 (4th ed. 1982) (“In mechanical testing, as in the field of mechanics in general, we attach rather specific meanings to certain terms, some of which are interpreted more loosely in everyday life.”).

23. WILLIAM ROSEN, *THE MOST POWERFUL IDEA IN THE WORLD: A STORY OF STEAM, INDUSTRY, AND INVENTION* 68 (2010).

24. HENRY PETROSKI, *TO FORGIVE DESIGN* 87 (2012) (also linking Rosen's observation to the Chinese proverb that “failure is the mother of success”); see also Gene I. Rochlin, *Informal Organisational Networking as a Crisis-Avoidance Strategy: US Naval Flight Operations as a Case Study*, 3 *INDUS.*

neer's preoccupation with failure enhances an understanding of the causal environments within which structures, products, or systems operate:

In all cases of surprise or failure, the greater technological tragedy is not having failures but not learning the correct lessons from them. Every failure is a revelation of ignorance, an accidental experiment, a found set of data that contains clues that point back to causes and further back to mistakes that might have been made in design, manufacture, and use. Not to follow the trail to its source is to abandon an opportunity to understand better the nature of the technology and our interaction with it.²⁵

For this reason, although engineers “are always striving for success,” it is nevertheless “imperative that the realistic prospect of failure be kept in the forefront of every engineer's mind.”²⁶ As a result of this imperative, engineers are said to be “always interested in the worst-case scenario.”²⁷

In the twentieth century, many branches of engineering witnessed a methodological drift away from direct experimental testing and towards theory and computer-based models.²⁸ This shift to analytical models and away from direct experimentation prompted intramural conflicts between those engineers who preferred to focus on their computationally sophisticated model simulations and those who emphasized instead the need for experimental data to verify the model and, more importantly, understand better the causal environment—or, in other

CRISIS QUART. 159, 170 (1989) (“It is said in the Navy that every lesson is ‘written in the blood of failures.’”).

25. PETROSKI, *supra* note 24, at 45.

26. *Id.* at 175, 198 (also noting that “failure is seldom far from [engineers'] minds”).

27. IZURU TAKEWAKI ET AL., IMPROVING THE EARTHQUAKE RESILIENCE OF BUILDINGS: THE WORST CASE APPROACH 1 (2013).

28. See, e.g., JOSEF SINGER ET AL., BUCKLING EXPERIMENTS: EXPERIMENTAL METHODS IN BUCKLING OF THIN-WALLED STRUCTURES 3 (1998); Wolfgang G. Knauss, *Perspectives in Experimental Solid Mechanics*, 37 INT'L J. SOLIDS & STRUCTURES 251, 256 (2000). Direct experimentation permits an analyst to determine with a high degree of certainty the stress distribution in a system or machine component in actual operation without needing to know with the same degree of certainty the exact nature of the forces acting on the part of system. See MIKLÓS HETÉNYI, HANDBOOK OF EXPERIMENTAL STRESS ANALYSIS, at v (1950). On the other hand, it sacrifices somewhat the ability to individuate the precise causal impact of each contributing force on the tested subject. See HANDBOOK OF EXPERIMENTAL SOLID MECHANICS, at v (William N. Sharpe, Jr. ed., 2008); HETÉNYI, *supra* at v. Where it is impossible to experiment directly on a test subject, a stress analyst will often construct a smaller model and conduct tests based on assumptions regarding the relationship between the scaled model and the full sized construction it represents. See HENDRY, *supra* note 20, at 1.

words, the relevant model parameters—on which the model's structure should depend.²⁹ These debates presaged later conflicts in other fields, including finance, about the relative virtues and shortcomings of computer-driven analytical models and direct experiments. The point here is that engineering, with its demands to understand how causes can trigger failure in the real world, could not afford to sidestep this debate. Analytical theory, along with the computer technologies that facilitated its rise, came to co-exist with hard-nosed experimentation.³⁰

One particular analytical modeling method—the so-called Monte Carlo method—is used extensively by engineers and financial risk managers, and accordingly bears special mention in this context.³¹ Monte Carlo methods model a system or structure mathematically, but randomly generate values in a series of simulations for the relevant variables that affect the outputs of interests to the analyst.³² As a result, Monte Carlo simula-

29. See SINGER ET AL., *supra* note 28, at 5 (“New phenomena have still to be found and properly understood in physical tests, before even the powerful computers of today can give a reliable simulation and then extend the range of parameters.”); Bruce G. Johnston, *Buckling Behavior Above the Tangent Modulus Load*, 128 TRANSACTIONS AM. SOC'Y CIVIL ENG'RS 819, 820 (1963) (“There are many advantages in simulated tests, carried out with the aid of a computer, in comparison with real tests in an actual testing machine. No machining is involved, no materials need be acquired, and there is no scatter in the test results.”); Knauss, *supra* note 28, at 256–57 (comparing over-reliance on analytical computer models over experimental methods to “aspects of the scholasticism during the 13th century, which fathered the far reaching philosophical innovations by Roger Bacon”).

30. See HUGH W. COLEMAN & W. GLENN STEELE, EXPERIMENTATION, VALIDATION, AND UNCERTAINTY ANALYSIS FOR ENGINEERS 2 (3d ed. 2009) (“Experimental information is almost always required at one or more stages of the [analytical] solution process, even when an analytical approach is used. Sometimes experimental results are necessary before realistic assumptions and idealizations can be made so that a mathematical model of the real-world process can be formulated using the basic laws of physics. In addition, experimentally determined information is generally present in the form of physical property values and the auxiliary equations . . . necessary for obtaining a solution.”).

31. Monte Carlo simulations are so named because the method of generating random numbers resembles a “roulette-like machine of the kind utilized in the gambling Casinos of the Monte Carlo Principate.” ENRICO ZIO, THE MONTE CARLO SIMULATION METHOD FOR SYSTEM RELIABILITY AND RISK ANALYSIS 1 (2013). A late nineteenth century mathematician named John William Strutt Rayleigh invented basic Monte Carlo modeling techniques under the name “random walks.” See DIMITRIS N. CHORAFAS, RISK MANAGEMENT TECHNOLOGY IN FINANCIAL SERVICES: RISK CONTROL, STRESS TESTING, MODELS, AND IT SYSTEMS AND STRUCTURES 145 (2007).

32. See DOUGLAS W. HUBBARD, THE FAILURE OF RISK MANAGEMENT: WHY

tion models generate probabilistic assessments of system failures even before any particular failure has occurred as an operational matter. Recalling Petroski's maxim that engineers learn from failure,³³ Monte Carlo methods permit engineers to simulate failures with computers and learn from them before they ever happen.

B. SYSTEMS RELIABILITY ANALYSIS

While stress analysis has its roots in mechanical and structural engineering, its utility became apparent to other engineers as well. In the second half of the twentieth century, engineers developed a new branch of their discipline known as "systems reliability analysis" (SRA). SRA is inspired by the concept of *reliability*, which is tied to the concept of failure that undergirds all stress analysis. When used in this context, reliability is a statistical concept describing the probability that a system will not experience failure during a specified time period under given operating conditions.³⁴ The SRA engineer's job is to identify potential sources of failure and design the system to avoid those failures.³⁵ SRA engineers utilize visual diagrams and logic models known as *fault trees* and *event trees* to assist them in modeling the ways in which system failures can occur, particularly when engineering complicated systems such as nuclear power plants, transportation infrastructure systems, missile defense systems, resource extraction projects, space travel expeditions, and robotics.³⁶ SRA can be seen as a precursor to stress testing in the financial regulatory context because it contemplates *human judgment* as a source of system failure.³⁷ In this respect it is more readily adopted as a prototype for regu-

IT'S BROKEN AND HOW TO FIX IT 60–63 (2007).

33. See *supra* note 24 and accompanying text.

34. See DEP'T OF DEF., HANDBOOK ON RELIABILITY GROWTH MANAGEMENT 2 (2011); BALBIR S. DHILLON, RELIABILITY ENGINEERING IN SYSTEMS DESIGN AND OPERATION 5 (1983); ERNST G. FRANKEL, SYSTEMS RELIABILITY AND RISK ANALYSIS 11–12 (1988); PATRICK D.T. O'CONNOR, RELIABILITY ENGINEERING 1 (1988).

35. MARVIN RAUSAND & ARNLJOT HØYLAND, SYSTEM RELIABILITY THEORY: MODELS, STATISTICAL METHODS, AND APPLICATIONS 73 (2004).

36. See RAUSAND & HØYLAND, *supra* note 35, at 96–124; Clifton A. Ericson II, *Fault Tree Analysis—A History*, in PROCEEDINGS OF THE 17TH INT'L SYS. SAFETY CONF. 1 (1999), available at <http://www.fault-tree.net/papers/ericson-fta-history.pdf> ("The fundamental concept of Fault Tree Analysis is the translation of the failure behavior of a physical system into a visual diagram and logic model.").

37. See RAUSAND & HØYLAND, *supra* note 35, at 69.

lating corporate entities.

Though materials stress testing and SRA methods trace their origins to the engineering discipline, the concept is by no means limited to that context; after all, perhaps the most widely recognized stress testing application comes from cardiology, where a stress test consists of induced cardiovascular stress through exercise or pharmacological agents followed by imaging. The stress test's purpose is to progressively overload the cardiovascular system in order to reveal abnormalities not present at rest.³⁸ It is a “focal point in the diagnosis and prognosis of cardiovascular disease.”³⁹ And cardiology is just one salient example. The relevant scope of stress testing as a technique is as wide as the range of structures, materials, or systems that impact any matter of societal import—from the cardiovascular health of a single individual to citywide transportation systems and power plants. And stress analysis also applies, as explained below, to the financial system and the institutions that transact in it.

II. HOW DO STRESS TESTS WORK IN FINANCE?

A. WHY STRESS IS A PROMISING REGULATORY TOOL

The paramount concern of financial regulators is *financial failure*—i.e., the insolvency of an institution or the inability of a financial market to perform reliably and consistently.⁴⁰ Regulators are charged with curbing failure on both institutional and systemic levels.⁴¹ Their supervisory duties consist primarily of minimizing institutional failures in order to limit recourse to government guarantees (e.g., deposit insurance, state insurance guaranty funds) that are only partially recouped through ex post assessments on other institutions. Their system-wide regulatory duties consist primarily of preventing the break-

38. ROBERT S. ELIOT, *STRESS AND THE HEART: MECHANISMS, MEASUREMENTS, MANAGEMENT* 105 (1988).

39. David Akinpelu, *Treadmill Stress Testing*, MEDSCAPE, <http://emedicine.medscape.com/article/1827089-overview> (last visited Apr. 21, 2014).

40. See Jean Pierre Sabourin, *The Deposit Insurer's Role in Maintaining Financial Stability*, in *SYSTEMIC FINANCIAL CRISES: RESOLVING LARGE BANK INSOLVENCIES* 59, 59 (Douglas D. Evanoff & George G. Kaufman eds., 2005) (“Dealing effectively with systemic financial crises and the resolution of large bank insolvencies has always been an important subject for those working in the financial system safety net. But, it is becoming even more critical in a world of ever greater consolidation and globalization in financial services.”).

41. *Id.* at 61.

down in key financial functions such as the payments system, the monetary policy transmission mechanism, and savings transformation. A separate, but related danger against which regulators guard is the possibility that institutional or market failures ramify throughout the interconnected financial system, causing further failures with attendant system instabilities and losses to government guarantees.

Outside of an actually experienced failure, stress testing practices comprise the settings where corporate managers and boards engage most directly with the concept of failure. On account of that preoccupation with failure, it is hardly surprising that regulators have adopted stress tests and stress analysis as regulatory tools. If anything, it is surprising that it took them so long. Nuclear engineers conduct stress tests and other SRA-type exercises in order to ensure (1) the reliable delivery of electricity, on which businesses, households, and transportation systems depend; and (2) the avoidance of accidents, on which the environment depends.⁴² Financial institutions play analogous roles and present similar risks in the economic context. Their critical roles in facilitating payment systems, savings intermediation, and monetary policy transmission comprise the grid through which finance flows and commerce is enabled. And the consequences of a financial crisis can devastate the wider economy, much like the catastrophic effects of a nuclear reactor incident on the surrounding environment.

B. STRESS TESTS ARE A KEY COMPONENT OF EXISTING RISK MANAGEMENT SYSTEMS

Over the past three decades, stress testing techniques have become an integral part of the risk management infrastructure of a typical financial firm. In the late 1980s and 1990s, most financial firms developed systems of risk management to respond to increased exposures to market, credit, and operational risks on account of the proliferation of derivatives during the same period.⁴³ From a corporate governance perspective, this revolution in risk management reflected the board of director's

42. Cf. *Safety of Nuclear Power Reactors*, WORLD NUCLEAR ASS'N, <http://www.world-nuclear.org/info/Safety-and-Security/Safety-of-Plants/Safety-of-Nuclear-Power-Reactors/> (last visited Apr. 21, 2014) (discussing the implementation of stress tests to prevent accidents, as well as the importance of reliability).

43. See Robert F. Weber, *An Alternative Story of the Law and Regulation of Risk Management*, 15 U. PA. J. BUS. L. 1005, 1020 (2013).

responsibility to set corporate risk policy and management's delegated responsibility to put that policy into action.⁴⁴ Though the initial impetus to develop risk management systems originated in the private sector, public bank regulators quickly perceived that their ability to achieve their statutory mandates of institution-level safety and soundness and system-wide financial stability increasingly came to depend on the effectiveness of those systems. Accordingly, one of the key themes of bank supervision from the late 1980s to today is the redirection of supervisory resources towards fostering robust risk management systems within banks.⁴⁵ The use of stress testing as a regulatory technique forms part of this effort.⁴⁶

1. Brief Introduction to Risk Management Systems

At its most fundamental level, risk management is about identifying bad outcomes that could occur in an uncertain future and taking deliberate action to shift the odds in a firm's favor.⁴⁷ Risk management thus comprises risk *assessment* and risk *control*.⁴⁸ The risk assessment process requires considera-

44. *Id.* at 1044.

45. See *Lessons Learned in Risk Management Oversight at Federal Financial Regulators: Hearing Before the Subcomm. on Securities, Insurance, and Investment of the S. Comm. on Banking, Housing, and Urban Affairs*, 111th Cong. 9 (2009) (statement of Timothy W. Long, Senior Deputy Comptroller, Bank Supervision Policy and Chief National Bank Examiner, Office of the Comptroller of the Currency), [hereinafter Long Statement], available at <http://www.gpo.gov/fdsys/pkg/CHRG-111shrg52966/pdf/CHRG-111shrg52966.pdf> ("Risk management is a key focus of our large bank supervision program."); DANIEL K. TARULLO, *BANKING ON BASEL: THE FUTURE OF INTERNATIONAL FINANCIAL REGULATION* 274 (2008) (noting that because the "risks associated with the complexity and pace of large bank activities cannot be effectively contained even with sophisticated rules . . . the emphasis increasingly has been on fostering robust risk management systems within the banks themselves"); Ben Bernanke, Chairman, Bd. of Govs. of the Fed. Reserve Sys., Remarks at the Stonier Graduate School of Banking, Washington, D.C. (June 12, 2006), available at <http://www.federalreserve.gov/newsevents/speech/bernanke20060612a.htm> (characterizing regulators' "assessment of the quality of a bank's procedures for evaluating, monitoring, and managing risk" as the "heart of the modern bank examination").

46. See BASEL COMM. ON BANKING SUPERVISION, *PRINCIPLES FOR SOUND STRESS TESTING PRACTICES AND SUPERVISION* 5 (2009) [hereinafter BCBS, *STRESS TESTING PRINCIPLES*], available at <http://www.bis.org/publ/bcbs155.pdf>.

47. See DAN BORGE, *THE BOOK OF RISK* 4 (2001).

48. Cf. PHILIPPE JORION, *VALUE AT RISK: THE NEW BENCHMARK FOR MANAGING FINANCIAL RISK* 3 (2d ed. 2001) ("*Risk management* is the process by which various risk exposures are identified, measured, and controlled.").

tion of a firm's objectives.⁴⁹ The risk assessment process requires consideration of a firm's objectives.⁵⁰ A risk to be controlled, and therefore "managed," is one that impacts the achievement of those objectives. One risk manager explains that the "Holy Grail of risk management" is "to find the best possible decision to make when faced with uncertainty."⁵¹

This conception of risk presents a subtle contrast with other definitions of risk that use risk as a measure of volatility alone.⁵² While the volatility of asset returns and interest rates is obviously relevant to a financial institution, I argue that risk management, properly conceived, should not equate risk with volatility alone. The idea that risk connotes volatility dates from 1921, when economist Frank Knight distinguished uncertainty from risk on the grounds that latter was measurable and the former was not.⁵³ This Knightian definition of risk has an impressive legacy, but it is inadequate to describe risk management for two reasons.

First, it is too broad because it includes all uncertain future events that are capable of probabilistic measurement. Risk management, however, is only concerned with uncertainty to the extent it might impact organizational objectives.⁵⁴ It would be absurd to talk about the meanderings of an ant as risky until someone wagered a bet on the next movement. Second, it is in another respect too narrow because risk management is concerned with unmeasurable uncertainty as well. In finance it is a commonplace to use risk to refer to the variance of returns from the expected return—a definition that is measured quan-

49. See David Garland, *The Rise of Risk*, in RISK AND MORALITY 48, 50–51 (Richard V. Ericson & Aaron Doyle eds., 2003); Ian Hacking, *Risk and Dirt*, in RISK AND MORALITY, *supra* at 21, 22.

50. See Garland, *supra* note 49, at 50–51; Hacking, *supra* note 49, at 22.

51. BERGE, *supra* note 47, at 12.

52. Risk and uncertainty are contested terms even within the risk management field. HUBBARD, *supra* note 32, at 79 ("Concepts about risk and even the word *risk* are a source of considerable confusion even among those who specialize in the topic.")

53. See FRANK H. KNIGHT, RISK, UNCERTAINTY, AND PROFIT 233 (1921) ("To preserve the distinction between the measurable uncertainty and an unmeasurable one we may use the term 'risk' to designate the former and the term 'uncertainty' for the latter.")

54. See HUBBARD, *supra* note 32, at 81–84; Glyn A. Holton, *Defining Risk*, 60(6) FIN. ANALYSTS J. 19, 21 (2004) ("According to common usage, risk entails both uncertainty and exposure—possible consequences. Knight's distinction addresses only the uncertainty.")

titatively.⁵⁵ In her revealing research performed while embedded in bank risk management departments, Anette Mikes found evidence of two prevailing “calculative culture” types within risk management departments: quantitative enthusiasm and quantitative skepticism.⁵⁶ Mikes found that while all the risk management departments she examined made extensive use of quantitative information and statistical techniques, the quantitative skeptics considered risk measurements only as trend indicators to be taken into consideration alongside other qualitative criteria such as “managerial discretion, experience and judgment.”⁵⁷ Quantitative skeptics regard quantitative risk measurement tools as “learning machines.”⁵⁸ And yet a Knightian definition of risk would fail to include such information.

2. Brief Introduction to Value-at-Risk, a Building Block of Risk Management Systems

The idea of value-at-risk (VaR) connects the general concept of risk management to the concrete practices of stress testing at financial institutions. VaR models can be thought of as the basic building blocks of risk measurement and they also comprise the operational settings within which most financial stress tests are conducted.⁵⁹ VaR is an expression of the amount of possible loss to which a financial firm is exposed.⁶⁰ One risk management commentator refers to VaR as the “holy grail of

55. See Robert T. Miller, *Oversight Liability for Risk-Management Failures at Financial Firms*, 84 S. CAL. L. REV. 47, 59–60 (2010).

56. Anette Mikes, *Risk Management and Calculative Cultures*, 20(1) MGM'T ACCOUNTING RES. 18, 22 (2009).

57. Anette Mikes, *Chief Risk Officers at Crunch Time: Compliance Champions or Business Partners*, 2 J. RISK MGM'T FIN. INSTITUTIONS 7, 15 (2008).

58. Mikes, *supra* note 56, at 27.

59. COMM. ON THE GLOBAL FIN. SYS., BANK FOR INT'L SETTLEMENTS, STRESS TESTING AT MAJOR FINANCIAL INSTITUTIONS: SURVEY RESULTS AND PRACTICE 3 (2005) [hereinafter BIS, STRESS TESTING PAPER]. *But see* Robert F. Weber, *The Corporate Governance Case for Deliberation-Oriented Stress Testing Regulation*, 39 J. CORP. L. (forthcoming 2014) (describing “expected shortfall” risk measurement techniques as an alternative to VaR).

60. A comprehensive definition of VaR would disaggregate the term into the most direct meaning expressed above in the text and also the VaR “procedures” by which the VaR amount is generated, the statistical VaR “methods” utilized in the VaR procedures employed by the firm, and the VaR “approach” to risk management with its concomitant enthusiasm for risk quantification. *See* KEVIN DOWD, BEYOND VALUE AT RISK: THE NEW SCIENCE OF RISK MANAGEMENT 21 (1998).

risk management.”⁶¹ From the inception of VaR methods, risk managers self-consciously designed VaR methods in a manner to facilitate a new deliberative discourse among corporate stakeholder on how risk permeates across a financial firm: “Subject to the simplifying assumptions used in its calculation, [VaR] aggregates all of the risks in a portfolio into a single number suitable for used in the boardroom, reporting to regulators, or disclosure in an annual report.”⁶² Some risk management commentators saw in VaR the potential for an aspirational overhaul of corporate governance to a new rationalized, quantitative approach to managing not only risk but the entire firm.⁶³

Risk managers express VaR as the maximum expected loss a firm will face within a specified probability level (known as the “confidence level”) over a particular time period (known as the “time horizon”).⁶⁴ The VaR model allows risk managers to make the following statement: “We are *X* percent certain that we will not lose more than *V* dollars in the next *N* days.”⁶⁵ The most common variant of VaR methodology by far is historical simulation,⁶⁶ which consists of a computational phase and an interpretive-planning phase.

In the computational phase, the firm takes its current assets and liabilities and revalues them on the basis of daily historical values for relevant risk factors (e.g., interest rates, exchange rates) and prices over a predetermined look-back period, most typically 500 days.⁶⁷ This is essentially a counter-

61. LINDA ALLEN ET AL., UNDERSTANDING MARKET, CREDIT, AND OPERATIONAL RISK: THE VALUE AT RISK APPROACH 3 (2004) (“[M]arket forces during the late 1990s created conditions that led to the evolution of VaR as a dominant risk measurement tool for financial firms.”); JORION, *supra* note 48, at xxi; *cf.* note 51 and accompanying text.

62. Thomas J. Linsmeier & Neil D. Pearson, *Value at Risk*, 56 FIN. ANAL. J. 47, 48 (2000).

63. *See* DOWD, *supra* note 60, at 20–22.

64. CROUHY ET AL., *supra* note 6, at 187; DOWD, *supra* note 60, at 39.

65. JOHN C. HULL, RISK MANAGEMENT AND FINANCIAL INSTITUTIONS 196 (2007).

66. *See* Christophe Pérignon & Daniel R. Smith, *The Level and Quality of Value-at-Risk Disclosure by Commercial Banks*, 34 J. BANKING & FIN. 362, 367–70 (2010) (finding that 73% of large international banks that disclose their VaR methodologies, a group that itself comprises 65% of all such banks, use historical simulation methods).

67. *See* JAMES GLEASON, RISK 187 (2000) (“In a nutshell, you run your daily [profit and loss] process over and over again, using the history of past daily market price moves.”). For assets traded on markets, risk managers can simply plug in the appropriate historical prices. For untraded assets, risk

factual exercise: “assuming we had this portfolio yesterday, and the day before yesterday, and so forth, what would our net profit-loss position have been?”⁶⁸ When the firm has valued its assets and liabilities over the relevant time horizons comprising the look-back period, it ranks the net profit-loss outcomes from worst to best.⁶⁹

In the interpretive-planning phase, risk managers decide on the applicable confidence level. If the confidence level is, say, five percent, they will look to the fifth percentile of outcomes—thus, if the look-back period is 500 days then the fifth percentile would be the twenty-fifth worst net profit-loss outcome.⁷⁰ The difference between current market values of the portfolio and this fifth percentile is then interpreted as the “value” that is “at risk” over the next day.⁷¹ A higher confidence level will result in a higher VaR number—e.g., if the confidence level is 99% instead of 95% then the value at risk will be the difference between current market values and the *fifth* worst net profit-loss outcome.

As thus described, the term “value-at-risk” is a misnomer. Of course, the “value” that is “at risk,” as a layperson understands the term, is always the entire portfolio.⁷² In fact, VaR really tells you what you should expect to lose a few times a year.⁷³ One risk management text summarizes this point:

[VaR] is not the answer to the simple question: How much can I lose on my portfolio over a given period of time? The answer to this question is “everything,” or almost the entire value of the portfolio! . . . Instead, [VaR] offers a probability statement about the potential change

managers must construct a pricing model that specifies how the identified risk factors affect the asset’s value.

68. JORION, *supra* note 48, at 221 (observing that historical VaR methods do not “represent an actual portfolio but rather reconstruct[] the history of a hypothetical portfolio using the current position”). If the time horizon is longer—say, 10 days—then risk managers make appropriate adjustments. Specifically, they typically assume that the VaR estimate for a time horizon consisting of N days is equal to the product of the 1-day VaR estimate times \sqrt{N} . See ALLEN ET AL., *supra* note 61, at 7 (referring to this method as the “square root rule”); HULL, *supra* note 65, at 203.

69. See Miller, *supra* note 55, at 63.

70. See DOWD, *supra* note 60, at 39.

71. See *id.*

72. See *id.* at 40 (noting that such a measure “is not very informative”).

73. See Jón Daniélsson, *The Emperor Has No Clothes: Limits to Risk Modeling*, 26 J. BANKING & FIN. 1273, 1275 (2002) (noting that VaR “is only concerned with the 99% loss level, or a loss which happens 2.5 times a year, implying that VaR violations have very little relevance to the probability of bankruptcy, financial crashes, or systemic failures”).

in the value of a portfolio resulting from a change in market factors, over a specified period of time. [VaR] is the answer to the following question: What is the maximum loss over a given period such that there is a low probability, say a 1 percent probability, that the actual loss over the given period will be larger?⁷⁴

VaR techniques are a linchpin of all risk management systems, whether the departments operating those systems are staffed with quantitative enthusiasts or quantitative skeptics.

Even putting aside the inherent arbitrariness with selecting a time horizon and confidence level, there are technical limitations of the VaR techniques themselves. These limitations are well-documented and, at the risk of venturing into well-trodden territory—particularly after the recent financial crisis—three brief points regarding VaR's limitations are helpful in order to better understand the utility of stress testing. These limitations also emerge in a new light when considered alongside the related debate within engineering concerning the relative advantages of direct experimentation and analytical modeling.⁷⁵

The first limitation is epistemic. The embedded assumption of all VaR techniques is that the uncertain future will resemble the past.⁷⁶ The point may be stated even more narrowly: the future will resemble a *non-random* sample of *recent* past conditions.⁷⁷ This point should not be over-emphasized; it is of course true that the past is a useful guide for predicting the future. But although VaR models almost always achieve their purpose in normal market conditions, they do not work well when it comes to the extreme, rare market movements that are most closely associated with failure.⁷⁸

74. CROUHY ET AL., *supra* note 6, at 187. Lore attributes the invention of the idea of VaR to risk managers within J.P. Morgan in the early 1990s. See GILLIAN TETT, FOOL'S GOLD: HOW THE BOLD DREAM OF A SMALL TRIBE AT J.P. MORGAN WAS CORRUPTED BY WALL STREET GREED AND UNLEASHED A CATASTROPHE 33–34 (2009). The term appeared for the first time in a publicly available source in an influential report on derivatives published by the Group of Thirty in 1993. See JORION, *supra* note 48, at 22.

75. See *supra* notes 28–37 and accompanying text.

76. See HULL, *supra* note 65, at 221 (“Additionally, we should bear in mind that historical simulation assumes that the joint distribution of daily changes in market variables is stationary through time. This is unlikely to be exactly true and creates additional uncertainty about the value of VaR.”).

77. See Miller, *supra* note 55, at 64–65. In some respects, weighting recent time series data improves the accuracy of VaR estimates due to the stickiness, or cyclicity, of volatility levels. See ALLEN ET AL., *supra* note 61, at 36.

78. CROUHY ET AL., *supra* note 6, at 162; PHILIPPE JORION, VALUE AT RISK: THE NEW BENCHMARK FOR MANAGING FINANCIAL RISK 357 (3d ed.

The second limitation is statistical. Even if we were convinced that the future would resemble the past, the look-back periods are typically too small to permit the reliable statistical inferences.⁷⁹ Adopting a longer look-back period increases the robustness of the data set but with fast-changing markets older data might reflect obsolete market conditions and therefore contribute little to the accuracy of the distribution.⁸⁰

Third, financial markets and institutions interact in complex ways. Even assuming that the values for relevant risk factors are normally distributed, in real financial markets a single anomaly can ramify quickly through a financial network, interacting with other risk factors in unpredictable ways and potentially increasing exposures dramatically.⁸¹ The models, by contrast, assume that financial markets resemble the weather and place the risk manager in the role of meteorologist. Like the molecules and energy that interact to produce weather phenomena, the millions of heterogeneous financial market participants ensure that their aggregate actions are susceptible to randomized model. While this assumption no doubt holds during most periods, it breaks down in crises in large part because stressed human actors are less predictable than stressed molecules.⁸²

Instead, the risk factors interact in ways that defy the ex ante assumptions of the model. Risk is therefore said to be *endogenous*, which leads to unpredictable embedded properties non susceptible to ex ante modeling. This might simply be another problem of statistical data insufficiency—risk managers just don't yet know how these interactions should be distributed as a statistical matter.⁸³ However, some question whether such complex phenomena as financial market crises are ever capable of being the subject of predictive modeling.⁸⁴

If the problem is simply one of data insufficiency, there is a partial solution: risk managers can build analytical models that generate profit-loss data based on assumptions about how relevant parameters (e.g., interest rates, stock prices) interact to

2007).

79. Miller, *supra* note 55, at 65.

80. See JORION, *supra* note 48, at 222.

81. See Danielsson, *supra* note 73, at 1276.

82. See *id.* at 1276–77.

83. See HUBBARD, *supra* note 32, at 185–87.

84. See Weber, *supra* note 10, at 691–97 (summarizing several experiments showing dramatic phase transitions in dynamically modeled financial markets even with simple computer-based agents).

produce historical returns.⁸⁵ The predominant analytical modeling method is the Monte Carlo method familiar from the earlier discussion of stress testing in engineering.⁸⁶ When risk managers use Monte Carlo methods, they generate repeated simulations of risk factors, each time re-valuing the portfolio according to the generated “scenario.”⁸⁷ They can then run thousands (or more) simulations to obtain a wider data set from which to draw conclusions about VaR values. The virtue of Monte Carlo is that it is not limited by the historical biases that hamper historical simulation. Instead of relying on the historical distribution of risk factors, risk managers have to design the stochastic process by which the factor values are generated.⁸⁸ This advantage also creates a model risk problem: risk managers need to model the distribution of risk factors (and also the correlations between them) in a manner that is untethered to observed historical facts.⁸⁹ Fortunately, risk managers can work around this problem by conducting multiple simulation runs with different sets of risk factor distribution models.⁹⁰ One of the major benefits of Monte Carlo simulation is that it offers the opportunity to engage in critical deliberation, including on failure possibilities, as risk managers construct risk factor models.⁹¹

C. A TAXONOMY OF FINANCIAL STRESS TESTING TECHNIQUES

The purpose of VaR, as noted above, is to communicate information concerning a firm’s vulnerability as it anticipates the

85. See HULL, *supra* note 65, at 233 (“The main alternative to the historical simulation approach for estimating VaR for market risk is what is known as the model-building approach . . .”); JORION, *supra* note 48, at 224–29. As recently as 2000 banks made heavy use of analytical models in what are known as the variance-covariance approach and the delta normal approach, though today such approaches have fallen into disuse. Pérignon & Smith, *supra* note 66, at 367 n.15. For a technical summary of these techniques, both of which require stylized statistical assumptions concerning the normal distribution of risk factor volatility, see DOWD, *supra* note 60, at 63–98; HULL, *supra* note 65, at 233–52.

86. See Pérignon & Smith, *supra* note 66, at 372 fig. 4 (finding that 21% of those large international banks that disclose their VaR methodologies, a group that itself comprises 65% of all such banks, use Monte Carlo methods); *supra* notes 31–33 and accompanying text.

87. CROUHY ET AL., *supra* note 6, at 212–16; *cf.* CHORAFAS, *supra* note 31, at 147 (noting that Monte Carlo “uses random sampling to determine some probabilistic property of a population of objects or events”).

88. JORION, *supra* note 48, at 224–26.

89. *See id.*

90. See CROUHY ET AL., *supra* note 6, at 216.

91. See CHORAFAS, *supra* note 31, at 155.

conditions within which it will operate in an uncertain future.⁹² The two dominant methods of VaR systems—historical simulation and Monte Carlo—accomplish that task by relying, respectively, on recent historical returns and assumptions concerning the behavior of, and relationships between, risk factors.⁹³ These assumptions make VaR easy to use and communicate within the firm.⁹⁴ Accordingly, they enhance a firm’s understanding of its vulnerabilities. But VaR still tells us very little about the underlying causes of failure and the channels through which it may emerge.⁹⁵ Their limitations ensure that VaR is only a part of a robust system of risk management.

Stress testing alerts management to adverse unexpected outcomes related to a variety of contingent circumstances and often provides an indication of the magnitude of the potential loss or funding gap.⁹⁶ The advantage of stress testing is its increased flexibility, when compared to VaR, to identify vulnerabilities to extreme events.⁹⁷ Stress tests are therefore the most direct connection between financial firm management and failure.

For historical and conceptual reasons, it is helpful to consider stress testing against the background of VaR techniques. Risk managers developed stress testing methodologies alongside VaR systems.⁹⁸ And several of the exercises that carry the “stress test” label are operationalized within VaR statistical models. As one risk management expert noted, “stress-testing means choosing scenarios that are costly and rare, and then putting them to a valuation model.”⁹⁹ While he overstates the case somewhat—for instance, war games are increasingly seen as a part of stress testing but are not susceptible to quantitative modeling—the links between stress testing and quantitative modeling techniques are historical and operational. This is

92. See *supra* notes 64–65 and accompanying text.

93. See CROUHY ET AL., *supra* note 6, at 206–16.

94. See *supra* notes 62–63 and related text.

95. Cf. CHORAFAS, *supra* note 31, at 155 (“To depend on facts without putting them under stress is most dangerous because facts alone say nothing about underlying causes.”).

96. See BCBS, STRESS TESTING PRINCIPLES, *supra* note 46, at 7.

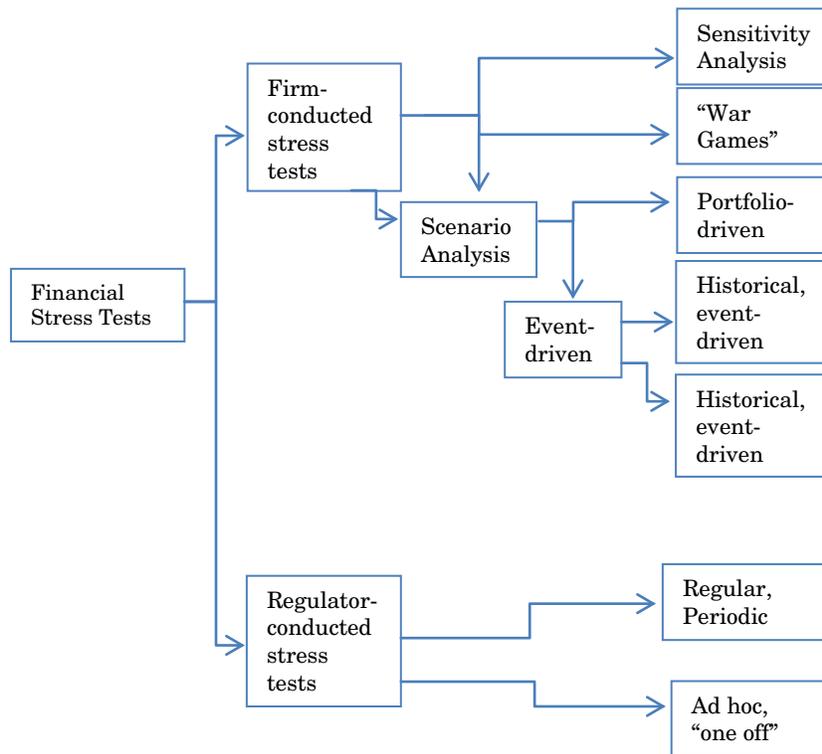
97. CROUHY ET AL., *supra* note 6, at 239–40.

98. See Drehmann, *supra* note 4, at 62 (“Notwithstanding that objectives are different for different stress tests current models share very much a common structure. . . . [T]his structure is rooted in the quantitative risk management framework.”).

99. Jeremy Berkowitz, *A Coherent Framework for Stress Testing*, 2 J. RISK 1, 2 (2000).

not to say that the uses of stress tests are cabined within VaR practices. When used properly, stress tests are “a complement, rather than a supplement,” to these statistical measures.¹⁰⁰

The discussion below develops a typology of the contexts within which stress testing has become an important component of risk management. Disparate methods of stress testing can be distinguished. An overarching distinction may be observed between *sensitivity* analysis and *scenario* analysis. Scenario analysis can take the form of *portfolio-driven* scenario analysis and *event-driven* scenario analysis. Event-driven scenario analysis may be further decomposed into *historical* scenario analysis and *hypothetical* scenario analysis.¹⁰¹ *War games* and *regulator-conducted* stress tests are addressed separately. The below schematic depicts these basic distinctions, which are further developed in this section:



100. BIS, STRESS TESTING PAPER, *supra* note 59, at 3.

101. See BCBS, STRESS TESTING PRINCIPLES, *supra* note 46, at 5.

1. Firm-Conducted Stress Tests

Firm-conducted stress tests can be divided into sensitivity analyses, scenario analyses, and direct “war game” scenarios. Brief descriptions of each category of exercise follow.

a. *Sensitivity Analysis*

When conducting sensitivity analysis, risk managers move isolated risk factors by a unit amount to determine how “sensitive” metrics such as profit and loss are to the changes.¹⁰² For example, the firm might consider the impact of a 100 basis point shift in bond yields, or a 10% decline in the stock market. Sometimes these tests are performed on single, isolated variables, and other times risk managers consider correlations between the stressed variable and other risk factors to determine how the firm would respond to a more comprehensive suite of changes in risk factors.¹⁰³

When conducting sensitivity analysis within the VaR context, risk managers posit values for one or more risk factors outside the scope of the VaR model.¹⁰⁴ For example, risk managers using historical simulation VaR might re-value the firm’s portfolio based on changes in interest rates that are outside the confidence interval or even outside the historical distribution entirely. The limitation of sensitivity analysis is that the stipulated changes in risk factor variables are unlikely to resemble a future real-life scenario.¹⁰⁵ Such tests nevertheless are very useful for verification purposes, allowing the firm “to form a first approximation of the impact on the firm of a move in a financial variable.”¹⁰⁶

b. *Portfolio-Driven Scenario Analysis*

Scenario analysis takes the further step of considering how specific real-world scenarios might cause those risk factor variables to shift.¹⁰⁷ When conducting portfolio-driven scenario analysis, risk managers examine an existing portfolio, discuss and individuate its vulnerabilities, and then imagine scenarios under which those vulnerabilities would become stressed.¹⁰⁸

102. See BIS, STRESS TESTING PAPER, *supra* note 59, at 8.

103. See *id.*

104. *Id.* at 4.

105. SHIRREFF, *supra* note 6, at 66–67.

106. BIS, STRESS TESTING PAPER, *supra* note 59, at 8.

107. *Id.* at 6.

108. *Id.*

The intellectual exercise is similar to the fault tree analysis that reliability engineers perform.¹⁰⁹ Some commentators refer to these obviously imaginative, highly counter-factual portfolio-driven tests as “reverse stress tests.”¹¹⁰ In effect, the risk managers attempt to reverse engineer a failure.

c. Event-Driven Scenario Analysis

By contrast, an event-driven stress test starts from an imagined scenario involving stressed risk factor variables in order to see how the firm or the portfolio might respond.¹¹¹ The scenario fixes the risk factor variables and subjects them to a risk model—usually the firm’s VaR model—with respect to a given set of exposures or portfolios, in the process generating a measure of the risk exposure to which that scenario corresponds.¹¹² It is therefore similar to sensitivity analysis, though it frequently considers multiple risk factor variables and is conceived of as a potentially real scenario, whether historical or hypothetical. The schematic below illustrates this process.



When conducting historical scenario analysis, risk managers attempt to reproduce the effects of extreme historical events.¹¹³ A hypothetical scenario is a significant market event that has no direct historical precedent. Obviously, the latter involves significantly greater exercise of managerial discretion and judgment in formulating the scenario.¹¹⁴ Risk managers frequently invite senior managers, front desk sales staff, and

109. See *supra* note 36 and accompanying text.

110. BCBS, STRESS TESTING PRINCIPLES, *supra* note 46, at 14.

111. BIS, STRESS TESTING PAPER, *supra* note 59, at 6.

112. See *id.* at 7.

113. See *id.* at 10 (highlighting the October 1987 stock market crash, the 1998 Russian debt default, the 2000 collapse of Long-Term Capital Management, the 1997 Asian financial crisis, and the 2001 U.S. terrorist attacks as frequently used scenarios); SHIRREFF, *supra* note 6, at 66.

114. See Berkowitz, *supra* note 99, at 2–3 (“The problem is that choosing stress-test scenarios is by its very nature subjective.”).

economists to these discussions.¹¹⁵ At the enterprise level, this broader participation also goes some way to ensuring that cross-business correlations and interactions are taken into account.¹¹⁶ One commentator has disaggregated event-driven scenarios into four types of simulated shocks: (1) scenarios that are expected “to occur more frequently than historical observations suggest”; (2) scenarios that have never occurred, but which might occur; (3) scenarios reflecting “the possibility that statistical patterns could break down in some circumstances”; and (4) scenarios reflecting “structural breaks” that could occur in the future.¹¹⁷ Hypothetical scenarios are particularly helpful to test the robustness of assumptions underlying statistical models for financial assets with limited historical data concerning risk.¹¹⁸ If properly conducted, event-driven scenario analysis can be used as a learning tool that “raises as many questions as it answers.”¹¹⁹

But for all their evident promise, event-driven stress tests have not been implemented effectively. In a 2005 study, the Bank for International Settlements found that some firms have difficulty selecting “big picture” hypothetical scenarios.¹²⁰ A 2009 Government Accountability Office report on risk management failures singled out banks’ failure to create a scenario for the hardly remote possibility of a severe downturn in the economy.¹²¹

d. War Games

War games lie on the outward bound of the stress testing definitional category, in large part because they do not fit neatly with analytical modeling tools such as VaR. I use the term here to connote direct simulations of scenarios where firm personnel “act out” how they would react to the appearance of

115. BIS, STRESS TESTING PAPER, *supra* note 59, at 7.

116. U.S. GOV’T ACCOUNTABILITY OFFICE, GAO-09-499T, FINANCIAL REGULATION: REVIEW OF REGULATORS’ OVERSIGHT OF RISK MANAGEMENT SYSTEMS AT A LIMITED NUMBER OF LARGE, COMPLEX FINANCIAL INSTITUTIONS: TESTIMONY BEFORE THE SUBCOMM. ON SECS., INS., & INVS. OF THE S. COMM. ON BANKING, HOUS., & URBAN AFFAIRS, 111TH CONG. 22–24 (2009) [hereinafter *GAO Risk Management Oversight Assessment*], available at <http://www.gao.gov/products/GAO-09-499T>.

117. Drehmann, *supra* note 4, at 73–74.

118. BCBS, STRESS TESTING PRINCIPLES, *supra* note 46, at 9.

119. ALLEN ET AL., *supra* note 61, at 5.

120. BIS, STRESS TESTING PAPER, *supra* note 59, at 8.

121. *GAO Risk Management Oversight Assessment*, *supra* note 116, at 22.

stressed market variables. Despite their ill fit with quantitative models, war games contribute to managerial understanding of uncertainty and how failure might develop.¹²² The authors of a book on political risk laud war games scenarios as being motivated “to inspire creative problem solving and to spur managers to think about unthinkable outcomes.”¹²³ The same could be said with respect to financial risk.

Unsurprisingly, the roots of war games can be traced in military history.¹²⁴ Military planners run “hard gaming” simulations conducted with computer models alone, “soft gaming” simulations involving indoor simulations, and “field training exercises” involving a large number of personnel across even thousands of square miles.¹²⁵ As early as the 1970s, corporate firms began to conduct their own direct simulations of adverse events. Royal Dutch-Shell was an early pioneer of these techniques, even taking its executives through the possibility of a sharp rise in oil prices a year before the 1973 Arab oil embargo.¹²⁶ Even today, Royal Dutch-Shell is applauded for its political war gaming scenario analysis in connection with investment projects.¹²⁷ The Deloitte consultancy developed a War Game-style simulation that it labeled a “stress test” to gauge the adequacy of measures developed by businesses and those responsible for IT and transportation infrastructure decisions in advance of the 2012 summer Olympic Games in London.¹²⁸ In finance, some commentators have called for regular use of war

122. See THE PEW CHARITABLE TRUSTS, STANDARDS FOR RAPID RESOLUTION PLANS 10 (2011), available at http://www.pewtrusts.org/uploadedFiles/wwwpewtrustsorg/Reports/Economic_Mobility/Standards_Rapid_Resolution_Plans.pdf.

123. IAN BREMMER & PRESTON KEAT, THE FAT TAIL: THE POWER OF POLITICAL KNOWLEDGE FOR STRATEGIC INVESTING 26 (2009).

124. B. Sheppard & D. Slavin, *The Use of War Game Simulations for Business Strategies*, in REAL-TIME AND DELIBERATIVE DECISION MAKING 77, 77 (Igor Linkov et al. eds., 2008).

125. *Id.* at 79–80; see also SHIRREFF, *supra* note 6, at 108 (“The British armed forces play war games once a year. Two sides code-named ‘Red’ and ‘Blue’ slug it out in virtual warfare. For them the purpose is to test equipment, communications and soldiers without the expense and losses associated with a real war.”).

126. BREMMER & KEAT, *supra* note 123, at 26–27; SHIRREFF, *supra* note 6, at 106–07.

127. See BREMMER & KEAT, *supra* note 123, at 26–27.

128. Keith Weir, “Stress Test” Highlights Olympic Transport Fears, REUTERS (June 18, 2012), <http://uk.reuters.com/article/2012/06/17/uk-olympics-business-stress-idUKBRE85G0SB20120617>.

gaming in developing resolution plans—also known as “living wills” or “funeral plans.”¹²⁹

The unique advantage of war game simulations is their ability to “simulat[e] human imponderables and behaviors” as they might unfold as failure approaches.¹³⁰ To harness these unique advantages of war games, testers apply aspects of direct experimental stress testing techniques from mechanical and structural engineering to the human organizational realm.¹³¹ But in order for the war games to produce useful failure-related information, the human decision makers involved in the war game must remain open to the possibility of failure.¹³² Otherwise, the war game becomes less a learning exercise and more of a validation of prior commitments and strategies. In his account of the Midway Battle with U.S. Naval forces over the Pacific, Japanese General Mitsuo Fuchida—who led Japan in the successful Pearl Harbor raid—recounts how Rear Admiral Matome Ugaki overruled decisions by the umpires of Japanese central command war games.¹³³ Consequently, the results of the game suggested that the Japanese naval conquest of the Pacific would proceed much more smoothly than expected.¹³⁴ Military assets were allocated in part based on the outcomes of the war games with disastrous consequences for the Japanese forces, which had underestimated the riskiness of its strategy.¹³⁵

The conceptual transition from stress testing of physical systems such as suspension bridges to social systems such as financial markets is marked by the presence of strategic human actors. Because people think and plan strategically, and then make decisions based on their plans, they are much more difficult to model than physical system components.¹³⁶ A social system’s stability depends not only on the risks to which it is sub-

129. See THE PEW CHARITABLE TRUSTS, *supra* note 122, at 1 n.1.

130. Sheppard & Slavin, *supra* note 124, at 80.

131. For a discussion of direct experimental stress testing techniques in the engineering discipline, see *supra* notes 29–30 and accompanying text.

132. THE PEW CHARITABLE TRUSTS, *supra* note 122, at 6 (“Systemically significant activities should be identified and provisions made to maintain them through a failure.”).

133. MITSUO FUCHIDA & MASATAKE OKUMITA, MIDWAY: THE BATTLE THAT DOOMED JAPAN, THE JAPANESE NAVY’S STORY 124–26 (Clarke H. Kawakami & Roger Pineau eds., 1955).

134. *Id.* at 96.

135. *Id.* at 96–98.

136. See W. Brian Arthur, *Complexity and the Economy*, 284 SCIENCE 107, 107 (1999); J. Dooyne Farmer et al., *Is Economics the Next Physical Science?*, 58 PHYSICS TODAY 37, 39 (2005).

ject at any given point, but also on our perceptions of those risks.¹³⁷ While it is true that “the odds [of] a 100-year storm do not change because people think that such a storm has become more likely,”¹³⁸ the odds of a liquidity run are very much affected by subjective human perceptions of its likelihood. This presents the endogeneity of risk problem discussed earlier.¹³⁹ War gaming allows risk managers to become more familiar with these odds, or at least with the causal network out of which such an event may spring. Once a sufficient number of games have been conducted, it may be possible to use the data in statistical models to apply to other contexts.

2. Regulator-Conducted Stress Tests

Regulator-conducted stress tests are performed by the financial regulator, not the regulated financial firms. This category of stress tests includes isolated, one-off stress tests as well as more regular, structured programs of stress testing.¹⁴⁰ The subjects of regular, structured programs of stress tests can be individual institutions or entire financial systems.¹⁴¹ The Financial Stability Assessment Program (FSAP), established by the International Monetary Fund (IMF) in response to the Asian financial crisis of the late 1990s, is the most prominent example of a regular stress testing program of entire financial systems. Under the FSAP, IMF personnel conduct an in-depth analysis of an individual country’s financial system.¹⁴² Each FSAP concludes with the preparation of a Financial System Stability Assessment (FSSA), which focuses on issues relevant to IMF surveillance.¹⁴³ One of the main components of the

137. JOHN KAMBHU ET AL., NAT’L RES. COUNCIL, NEW DIRECTIONS FOR UNDERSTANDING SYSTEMIC RISK 65 (2007).

138. *Id.* at 48.

139. *See supra* Part II.B.

140. *See* Marco Sorge, *Stress-Testing Financial Systems: An Overview of Current Methodologies* 1 (Bank for Int’l Settlements, Working Paper No. 165, 2004), available at <http://www.bis.org/publ/work165.pdf> (labeling these tests the “piecewise approach” and “integrated approach,” respectively).

141. *Id.* at 6.

142. FSAP assessments take place on a voluntary basis. As of 2009, over three-fourths of IMF member countries have undertaken an FSAP assessment. Public Information Notice No. 04/33, Int’l Monetary Fund, *IMF Executive Board Reviews Experience with the Financial Sector Assessment Program, Options for the Future, and Complementary Reforms in Surveillance and the Assessment of Standards and Codes* (Sept. 29, 2009), available at <http://www.imf.org/external/np/sec/pn/2009/pn09123.htm>.

143. *Id.*

FSSA is a system-wide stress test to assess financial stability.¹⁴⁴ Federal Deposit Insurance Corporation (FDIC) economist Paul Kupiec describes the FSAP stress tests as follows:

Stress tests often are designed with the cooperation of the national authorities with a goal to quantify the systemwide exposures that may result from a significant change in financial market fundamentals. The scenarios are usually specified as partial equilibrium “what if” type exercises where estimates of the balance sheet consequences of a financial event are quantified using data from a number of institutions or with aggregate data.¹⁴⁵

Kupiec judges the stress tests to be of little informational value due to regulators’ lack of access to information on banks’ off-balance sheet and derivative positions.¹⁴⁶ He also laments the partial equilibrium assumptions of the stress tests—that is, their focus on a single factor in isolation without considering how stress in that factor might interact dynamically with other risk factors.¹⁴⁷

In recent years, policymakers have focused increasingly on regulator-conducted stress tests. U.S. bank regulators conducted a one-off, ad hoc round of stress tests for the nineteen largest bank holding companies in 2009 as a linchpin of their efforts to stabilize the reeling banking industry.¹⁴⁸ Since then, European banking authorities have conducted several rounds of ad hoc stress tests of banks subject to their supervision.¹⁴⁹ Furthermore, the Dodd-Frank Wall Street Reform and Consumer Protection Act of 2010 made regular, periodic stress testing of banks a new responsibility for bank regulators.¹⁵⁰ These matters are taken up in greater detail below in Part IV.C.

The below table lays depicts the rough taxonomy of stress tests presently in use at financial institutions. It highlights the extent of managerial involvement in the design and conduct of tests. It also registers the extent to which the tests rely on VaR, which, as discussed above, yields precise, quantifiable numerical outputs. Notice that managerial involvement in stress testing is highest where VaR usage does not predominate. The

144. See Paul Kupiec, *The IMF-World Bank Financial Sector Assessment Program: A View from the Inside*, in SYSTEMIC FINANCIAL CRISES: RESOLVING LARGE BANK INSOLVENCIES 69, 71 (Douglas D. Evanoff & George G. Kaufman eds., 2005).

145. *Id.*

146. *Id.* at 77.

147. *Id.*

148. See *infra* note 260 and accompanying text.

149. See *infra* notes 284–87 and accompanying text.

150. See *infra* note 265 and accompanying text.

trade-off between open-ended managerial involvement and the precision of largely automated analytical tools has significant implications for design of a regulatory system of stress testing. This point is taken up in the next Part.

Types of Stress Tests			Management Involvement	Usage of VaR	
Conducted by Regulator	Regular, Periodic	System-wide	None	Some	
		Institution-specific			
	Ad hoc, "One-off"	System-wide	None	Some	
		Institution-specific			
Conducted by Firm	Sensitivity	Single Isolated Variable	Low	Predominant	
		Multiple Variable	Low	Predominant	
	Scenario	Portfolio-Driven/ "Reverse Stress Tests"		High	No
		Event-Driven	Historical	Low	Some
			Hypothetical	High	Some
		"War Games"		High	No

III. ELABORATING THE DISTINCTION BETWEEN DELIBERATION-ORIENTED STRESS TESTING AND ASSURANCE-ORIENTED STRESS TESTING

Note that in all their manifestations highlighted above, stress tests share a common element: they provide information to decision makers concerning economic outcomes in an uncertain future. The category of "decision makers" includes, most prominently, risk managers, but also could include business managers, directors, regulators, investors, and even financial service consumers. This suggests a general point on risk management that has been underemphasized in both the technical and academic finance and law literature: namely, that VaR is only one of several modes by which information production occurs in financial firms. The role of risk managers is to adopt and administer analytical tools that generate information and then communicate it throughout the firm, at times competing for executive and board attention or triggering automated pro-

cesses—e.g., to sell off or hedge risk exposures.¹⁵¹ True to their roots in the engineering discipline, stress tests in finance are usually the predominant context in which *ex ante* information concerning failure and stress *that is not captured in the VaR framework* is produced. Therefore, the mix of tools that a particular firm or regulator adopts determines the decisional infrastructure within which a firm's managers and directors process and act on failure- and stress-related information. Each system of stress testing implies a determination as to how much institutional resources, imagination, and attention should be devoted to imagining "dark visions" as plausible scenarios that ultimately might color firm or regulator decision-making.¹⁵²

Of course, some uses of stress tests contribute more meaningfully than others to an understanding of the likely channels through which failure might emerge. The diversity of contexts in which stress tests can be put to use highlights the need for policymakers and scholars to make appropriate distinctions among types of stress tests.¹⁵³ This Article argues in favor of an overarching distinction between two broad stress testing approaches: *assurance-oriented stress testing* and *deliberation-oriented stress testing*. These terms describe the firm's *institutional orientation* to stress testing—i.e., its attitudinal settings and decisional infrastructure—but they are also helpful to characterize *regulatory* interventions according to the sort of institutional orientation regulators are seeking to promote. That is, a regulated bank may implement a stress testing program that draws heavily from an assurance-oriented stress testing approach, but the bank's regulators may intervene by promoting deliberation-oriented stress testing during bank examinations or through rulemaking. Thus, assurance-oriented stress testing and deliberation-oriented stress testing refer to the dispositions of the actors conducting and deliberating on

151. Cf. CROUHY, *supra* note 6, at xvii.

152. See Saul Hansell & Joseph B. Treaster, *The Job of Imagining the Unimaginable, and Bracing for It*, N.Y. TIMES, Oct. 20, 2001, at C1 (explaining how the September 11 attacks affected Wall Street risk management strategies).

153. See Drehmann, *supra* note 4, at 62 ("It is not enough to call for more stress testing but it is also important to know what should ultimately be achieved with the stress test results.").

the tests and the use to which they are put; the distinction does not categorize the particular types of tests themselves.¹⁵⁴

Stress testing as a regulatory tool is in its infancy, so the binary is admittedly rough, but certain generalizations can be observed from these two orientations. Each represents an encounter with the concept of failure, but from different perspectives. Assurance-oriented stress testing privileges static scenarios, draws from audit culture, relies on historical precedents, examines tested variables in isolation, requires precise estimation, and is motivated by compliance and verification concerns.¹⁵⁵ By contrast, deliberation-oriented stress testing privileges dynamic scenarios, draws from business operations culture, relies on imagination, considers the interactivity of tested variables, remains open to open-ended and uncertain outcomes, and is motivated by governance concerns.¹⁵⁶

A. ASSURANCE-ORIENTED STRESS TESTING

As with other audit-like exercises, assurance-oriented stress testing frequently involves communicative events for which the relevant audience carries expectations as to the format and precision of the communicated information.¹⁵⁷ Thus a program of regulator-conducted stress tests of banks with announced results to assuage market concerns about the banks' bad loans or securities would be an example of assurance-oriented stress testing. The test would be designed to communicate information to bank investors and counterparties, and even the broader economy, concerning the banks' susceptibility to failure. So too would a bank's internal sensitivity analysis¹⁵⁸ program designed to verify that the assumed distribution of risk factor variables in its VaR model did not miss sharp non-linear increases in expected losses immediately outside the distribution. Here, the audience is the risk management team itself and the exercise can be conceived of as a spot audit check of the VaR model's robustness. These expectations require a degree of precision in terms of the stress tests' final outputs. Greater precision facilitates communication to the audience, but it requires greater rigidity in the design of the stress test so

154. See *supra* Part II.C for a discussion of the various types of financial stress tests.

155. Weber, *supra* note 59, at 2 n.3.

156. *Id.*

157. On the importance of audience, see Drehmann, *supra* note 4, *passim*.

158. See *supra* notes 102–06 and accompanying text.

as to produce quantifiable metrics. This orientation towards stress testing, while undoubtedly of great importance, receives the most attention from industry and commentators, but the familiarity of these uses of stress tests disguises their technical and institutional limitations.

Because it requires greater precision, assurance-oriented stress testing does not readily accommodate qualitative business judgments or take into account the dynamism of financial markets.¹⁵⁹ Michel Crouhy makes this point with respect to the dynamic supply and demand of liquidity at financial institutions:

Market crises unfold over a period of time, during which [market] liquidity may dry out. Yet most scenario analyses are static in nature, i.e., are one-period models and do not allow for the trading of positions in an environment where liquidity varies from one period to the next [Such analyses] assume that events occur simultaneously, and that the portfolio [being tested] remains constant during the period. The modeling framework usually does not allow for dynamic hedging or the unwinding of positions.

. . . .

. . . .

. . . [T]raditional VaR models are, by construction, static. Increasing the risk horizon from one day to ten days, one month, or one year, does not make the model more dynamic. . . .

. . . .

Clearly, liquidity risk cannot be factored into this traditional static framework.¹⁶⁰

Crouhy is making a point about the endogeneity, or reflexive, nature of risk in financial markets.¹⁶¹ Assurance-oriented stress testing typically models an exogenous, or outside, shock to the test subject. The second-order (and third-order, and so forth) effects on the firm—to say nothing of those of counterparties and others transacting in the relevant market—are not modeled. This omission fails to take account of the financial system as it actually exists as a network of counterparties that interact and respond to one another.¹⁶² It echoes Kupiec's concerns

159. See BCBS, *STRESS TESTING PRINCIPLES*, *supra* note 46, at 2–3.

160. CROUHY ET AL., *supra* note 6, at 241.

161. See Drehmann, *supra* note 4, at 77–79; see also *supra* Part II.B.2's discussion of the technical limitations of statistical models.

162. See Prassana Gai & Sujit Kapadia, *Contagion in Financial Networks*, 466 PROCS. ROYAL SOC'Y A 2401 *passim* (2010); Erlend Nier et al., *Network Models and Financial Stability*, 31 J. ECON. DYNAMICS & CONTROL 2033 *passim* (2007).

about the partial equilibrium assumptions embedded in the IMF's FSAP stress tests.¹⁶³

The intellectual context of a stress test used as an assurance-verification tool takes the business decisions to allocate capital to a particular portfolio or business line as its starting point. The exercise of imagination by risk managers is not encouraged. The question posed—i.e., “What would be the loss if things got *this* bad?”—admits a single answer. Of course, this is also its virtue. Assurance-oriented stress testing only becomes a problem when it becomes the predominant feature of a firm's stress testing program rather than an ancillary feature of a broader, more deliberative program. Even a superficial comparison between (1) a spot-check sensitivity analysis addressing a change in prevailing interest rates and (2) a “reverse stress test”¹⁶⁴ exercise inquiring into how a portfolio might fail due to interest rate fluctuations illustrates the point. The former will yield a useable data point or a “pass-fail” or “yes-no” determination, whereas the latter would yield instead a set of further questions.¹⁶⁵ And it is easy to see how the precision of the result can lead to such stress tests being implemented to facilitate simple, binary audit-like tasks.¹⁶⁶

I do not mean to overstate these critiques of assurance-oriented stress testing. It contributes to risk managers' understanding of portfolio risk and its ease of use ensures frequent recourse to it. It is certainly more open-ended and flexible than closed VaR model systems. However, assurance-oriented stress testing on its own will not promote more ambitious regulatory objectives of financial stability and institution-level safety and soundness.¹⁶⁷ As such, attempts by policymakers, regulated financial institutions, or trade associations to cast assurance-oriented stress testing initiatives as something more than helpful verification tools should be evaluated skeptically. Such claims might reveal more about our anxiety concerning failure and stress more than the ability of assurance-oriented stress

163. See *supra* note 147 and accompanying text.

164. On reverse stress tests, see *supra* Part II.C.1.b.

165. See Nassim N. Taleb et al., *A New Heuristic Measure of Fragility and Tail Risks: Application to Stress Testing* 9–11 (Int'l Monetary Fund, Working Paper No. WP/12/216, 2012), available at <http://www.imf.org/external/pubs/ft/wp/2012/wp12216.pdf> (“[S]tress test results are often presented as having a binary ‘pass/fail’ outcome.”).

166. See Wessel, *supra* note 2, at A2 (discussing implications flowing from the fact that “the word ‘test’ . . . implies pass or fail”).

167. See *supra* Part III.A.

testing to contribute meaningfully to our ability to understand them. Michael Power has argued that a similar phenomenon exists with respect to risk management more generally: that the modern emphasis on risk management reflects our aspirations to control increasingly uncontrollable phenomena.¹⁶⁸

B. DELIBERATION-ORIENTED STRESS TESTING

These limitations of assurance-oriented stress testing emerge even more clearly when it is compared with deliberation-oriented stress testing. In contrast to assurance-oriented stress testing, deliberation-oriented stress testing privileges dynamic scenarios, draws from business operations culture, relies on imagination, considers the interactivity of tested variables, remains open to uncertainty, and is motivated by governance concerns. In this manifestation, stress testing emerges less as a mechanical tool for regulators and risk managers and more as an aspirational internal corporate governance norm that is preoccupied with stress, failure, and instability. Deliberation-oriented stress testing aspires to inform risk managers and other executives, as well as the board, of risks across the entire enterprise in addition to portfolio-specific risks. Furthermore, unlike assurance-oriented stress testing, it has no built-in audience; it is engaged in open-ended deliberation for

168. See MICHAEL POWER, ORGANIZED UNCERTAINTY: DESIGNING A WORLD OF RISK MANAGEMENT 5 (2007) (“[T]he phenomenon of ‘risk’ management of more and more aspects of social and organizational life reflects an increase in social *expectations* about the decidability and management of dangers and opportunities.”); see also LEE CLARKE, MISSION IMPROBABLE: USING FANTASY DOCUMENTS TO TAME DISASTER 4 (1999) (“[U]nder conditions of high uncertainty the promise and apparatus of rational planning itself becomes mainly *rhetorical*, becomes a means by which plans—independently of their functional relevance to the task—can be justified as reasonable promises that exigencies can be controlled. When uncertainty about key aspects of a task is high, rationalistic plans . . . become rationality badges, labels proclaiming that organizations and experts can control things that are, most likely, outside the range of their expertise. . . . Thus do organizations try to control the uncontrollable.”); Daniélsson, *supra* note 73, at 1274 (“There is . . . an increasing body of evidence that inherent limitations in risk modelling technology, coupled with imperfect regulatory design, acts more like a placebo rather than a scientifically proven preventer of crashes it is sometimes made out to be.”); Weber, *supra* note 43, at 1009 (presenting theory that risk management might be “a placeholder delimiting the range of objects that *demand organizational control* rather than the range of objects that are in fact *susceptible to such control*”).

the sake of better understanding and communicating failure vulnerabilities.¹⁶⁹

Andrew Haldane, Executive Director for Financial Stability at the Bank of England, has advanced what he calls a “five-point plan” to guide the supervision of stress testing practices at banks.¹⁷⁰ His proposals touch on several deliberation-oriented stress testing themes. First, he recommends both active involvement of regulators in crafting stress scenarios and the use of “reverse stress tests” by firm risk managers—what he refers to as “test[ing] . . . the destruction of their balance sheets.”¹⁷¹ Second, as a starting point, regulators should arrange for banks to conduct common stress scenarios to facilitate benchmarking of the results and the banks’ VaR models.¹⁷² Third, stress tests should be dynamic, so that the second and subsequent round interactions, and their consequences for system-wide risk, can be evaluated.¹⁷³ He describes such an approach as “iterative” stress testing and highlights the usefulness of such an approach to uncover phenomena relating to asset sales and liquidity hoarding that cannot be taken into account by assurance-oriented stress testing exercises.¹⁷⁴ Haldane describes such an exercise as a “stress test cum war game.”¹⁷⁵ Fourth, regulators must engage in a follow-up dialogue with bank management concerning the results of the stress tests.¹⁷⁶ Haldane’s fifth point is addressed to what regulators should do with the information gleaned from stress tests rather than how the tests should be conducted.¹⁷⁷ Taleb et al. similarly pick up on deliberation themes when they propose exploring the fragili-

169. See Drehmann, *supra* note 4, at 61 (observing that when used for purposes of communicating risk throughout a firm, stress tests should be “transparent and suitable for storytelling”).

170. Andrew G. Haldane, Exec. Dir., Bank of Eng., Why Banks Failed the Stress Test, Remarks at the Marcus-Evans Conference on Stress-Testing 8–9 (Feb. 9–10, 2009) [hereinafter Why Banks Failed the Stress Test], available at <http://www.bis.org/review/r090219d.pdf>.

171. *Id.* at 8.

172. *Id.*

173. *Id.* at 8–9.

174. *Id.*

175. *Id.* at 9.

176. *Id.*

177. Specifically, he argues that regulators should (1) take the results of the tests and information gleaned during post-test discussion into account when setting firms’ capital and liquidity buffers and (2) require public disclosure of results to enhance market discipline. *Id.* at 9.

ty of stress tests themselves—by asking “[h]ow would our estimates respond to a marginal change of the stress scenario?”¹⁷⁸

Even some industry representatives have come to advocate for a greater emphasis of deliberation-oriented stress testing, in spite of its greater costs.¹⁷⁹ The Counterparty Risk Management Policy Group III (CRMPG), the membership of which is comprised almost entirely of representatives of the largest internationally active banks,¹⁸⁰ published a series of recommendations on how to limit systemic risk in 2008.¹⁸¹ The CRMPG recommended several practices that resonate with deliberation-oriented stress testing, including: (1) disclosure to counterparties of stress test results on term sheets for difficult-to-understand financial instruments, which should facilitate learning;¹⁸² (2) reverse stress tests to address issues not captured by VaR models;¹⁸³ (3) incorporation of an “expanded suite of stress tests into a formalized production schedule, against which trends and developments in key risk factors and exposure amounts can be tracked”;¹⁸⁴ and (4) liquidity stress tests to gauge how unsecured and secured funding sources could dry up during times of stress.¹⁸⁵ The report also makes specific recommendations concerning the scenarios firms should run. Of particular note is consideration of: reputational and franchise-related risks, off-balance sheet exposures, and, as noted above, liquidity implications of stress (even on an intraday basis) at the consolidated- and entity-level.¹⁸⁶ This is a far cry from standard VaR-dominated risk management.

The great promise of deliberation-oriented stress testing is that it will foster a more mindful decisional infrastructure within financial institutions that institutionalizes the continuous deliberation on failure, catastrophe, and stress. By thus intervening into the corporate governance infrastructure of fi-

178. Taleb et al., *supra* note 165, at 6.

179. On the relative costliness of deliberation-oriented stress testing, see *infra* notes 447–56 and accompanying text.

180. See COUNTERPARTY RISK MGMT. POLICY GRP. III, CONTAINING SYSTEMIC RISK: THE ROAD TO REFORM, at vii–x (2008) [hereinafter COUNTERPARTY RM GROUP REPORT], available at <http://www.crmpolicygroup.org/docs/CRMPG-III.pdf>.

181. *Id.* at 18–19.

182. *Id.* at A-1 to A-2.

183. *Id.* at 84–85.

184. *Id.* at 85.

185. *Id.* at 27–28.

186. *Id.* at 28–30.

nancial institutions, stress testing mandates might curb managers' and boards' historic tendency to adopt excessively accommodating attitudes towards risk. In the process, it might facilitate regulatory efforts to promote the increasingly indeterminate and shifting statutory objective of financial stability—in other words, resilience to failure—under conditions of uncertainty, volatility, and authentic complexity.¹⁸⁷

C. FOUR HYPOTHETICAL CASE STUDIES OF STRESS TESTING

The four hypothetical case studies developed below illustrate both the wide breadth of techniques that go by the stress testing label and the contrast between assurance-oriented stress testing and deliberation-oriented stress testing. In the first and second case studies, two banks design and conduct stress tests to assess the likely impact of adverse developments in the leveraged loan business. The third and fourth case studies introduce the effects of regulatory influence on stress testing, each describing an initiative by state insurance regulators to use stress testing concepts in their supervision of life insurance companies.

1. “Bank of Hazard”

Consider two fictional firms—we'll call them Bank of Hazard (Hazard) and Bank of Prudence (Prudence)—each of which conducts a hypothetical event-driven stress test of its leveraged loan portfolio. Hazard submits its portfolio to a hypothetical scenario involving a rapid 250-basis point (bp)¹⁸⁸ jump in spreads on a leveraged loan index,¹⁸⁹ a scenario well outside the scope of the Hazard's VaR model. The risk managers selected

187. See generally Weber, *supra* note 10 (arguing that in many respects contemporary financial systems are characterized by complexity rather than mere complicatedness and discussing the implications).

188. One bp is one-hundredth of a percent. JOHN DOWNES & JORDON ELLIOT GOODMAN, *DICTIONARY OF FINANCE AND INVESTMENT TERMS* 56 (7th ed. 2006). Thus, a 400-bp increase in spreads is a 4.00% increase. A dramatic increase in leveraged loan spreads would signal the market's heightened perception of the riskiness of the loans and, *ceteris paribus*, would necessarily entail a corresponding price reduction in the market value of the loans.

189. The “spread” of a fixed income asset is the incremental amount of the asset's yield that is greater than a benchmark rate, such as LIBOR (the London Interbank Offered Rate). For example, if LIBOR is 2.5% and a leveraged loan yield is 7.0%, then its spread is 4.5%. A 250 bp increase in the spread would result in 7.75% yield. In this context, the “yield” expresses the ratio of contractual interest payments over the current market price of the loan. See *id.* at 803.

250 bp because it represented the “nightmare scenario” of the highest single-day jump in the index spread since its inception fifteen years ago. Hazard risk managers conduct the test by re-valuing each of the leveraged loans in the portfolio in light of the hypothetical spread increase and compute the aggregate value of the portfolio (\$450 million), which is \$100 million lower than its current market value (\$550 million). Next, they compare the \$100 million difference to the VaR figure based on a 95% confidence level, which is \$70 million—i.e., the VaR model suggests that Hazard should be 95% certain that it will not lose more than \$70 million over the applicable time horizon. The \$100 million loss is 133% of the VaR figure of \$75 million.

Because Hazard has a firm-wide risk limit policy that prohibits it from incurring any exposure for which a stress test reveals a potential loss of 150% of the VaR figure, the risk managers do not take any corrective action, but they send a memo to the leveraged loan department and the corporate treasury department notifying them that an abrupt 250-bp increase in spreads would very nearly bump up to the risk exposure limits set by top management. Some of the bankers take note, perhaps even pausing to consider which categories of leverage loans might be most susceptible to such a spread increase and whether similar profits—whether from interest on the loans or transaction fees—could be obtained without incurring that risk. Others silently whisper to themselves “There’s no way spreads are going to jump that high anytime soon; besides, my comp is based on deal volume and management told us to do more deals.”

2. “Bank of Prudence”

Risk managers at Bank of Prudence conduct a similar scenario test involving a 400 bp increase in the index spread. While the basic idea for the scenario originated during exchanges between bankers and risk managers within the leveraged loan department, the risk managers discussed the scenario with risk managers from the corporate treasury department. The corporate treasury risk managers point out that in such a scenario two other problems might occur: providers of Prudence’s short-term credit, such as repo lenders, will become concerned about (1) Prudence’s creditworthiness and (2) the value of the leveraged loans that Prudence has pledged as collateral for any secured short-term lending.

As such, they suggest the scenario should also test for the effects on Prudence's liquidity position of a simultaneous 25% increase in collateral demands from short-term funding sources. Another risk manager points out that such a dramatic increase in spreads and collateral demands might very well be accompanied by a failure of one of Prudence's larger counterparties that has been known to trade in the riskiest leveraged loans and rely heavily on overnight repo lending to fund its operations. The risk managers decide to run the base scenario with the 400 bp spread increase and the 25% increase in collateral demands and an alternative scenario that assumes that a large counterparty defaults on all its unsecured obligations to Prudence.

3. Isolated Stress Test of GMDB Exposures

Consider next two hypothetical regulatory examples of stress testing. In the first example, a state insurance regulator conducts an isolated, one-off sensitivity analysis of the guaranteed minimum death benefit (GMDB) variable annuity contracts for all the life insurance companies subject to its jurisdiction.¹⁹⁰ All of the tested companies had regulatory capital levels in excess of the "company action level" that would require regulatory intervention.¹⁹¹ Nevertheless, recent press coverage of the unfavorable experience of GMDB business in several companies domiciled in other jurisdictions has led to concerns about annuitants liquidating their GMDB contracts. The attenuated investor confidence concerns the state insurance regulator because large policy liquidations deplete an insurance company of assets to a point at which its claims-paying ability is impaired. This concern motivates the regulator to conduct the stress test, which assumes a stock market decline of 20%

190. GMDB contracts are a type of variable annuity contract. Whereas with a normal variable annuity the annuitant's contributions are invested in assets for which the annuitant and his or her beneficiaries bear the entire amount of market risk, with a GMDB contract, the insurance company guarantees a minimum payout in the event of the annuitant's death irrespective of the balance of the variable annuity account, thereby shifting a portion of the market risk from the annuitant's beneficiaries to the insurance company.

191. See Robert F. Weber, *Combating the Teleological Drift of Life Insurance Solvency Regulation: The Case for a Meta-Risk Management Approach to Principles-Based Reserving*, 8 BERKELEY BUS. L.J. 35, 61 n.103 (2011) (describing the risk-based capital regime and the role of "company action level" capital triggers).

over the course of three months.¹⁹² The regulator announces that it will require the holding company of any insurance company that “fails” the stress test to contribute additional capital to the company.¹⁹³ The regulator explains that a company will fail the stress test if its regulatory capital would be less than 200% of the “company action level” under the scenario. Of the ten insurance companies subjected to the stress test, nine pass and one fails. The failing company’s holding company is instructed to contribute an additional \$40 million of additional capital. The regulator issues a press release announcing the results of the stress test, which is reported in the industry press and sent via email to every registered insurance agent in the state.

4. Ongoing Program of Dialogic Stress Testing of GMDB Exposures

In the second example, the insurance regulator of another state with several large life insurance companies, similarly concerned by the volume of GMDB business underwritten by its domiciliary companies, establishes a stress testing program. The regulator publishes a circular instructing all insurance companies subject to its supervision to stress test their GMDB contract portfolios regularly for “equity market and related risks.” The circular provides for direct involvement of regulatory examiners in testing design and implementation, and regulatory review of results.

Several companies quickly contact their examiners and lament the lack of precision in the circular, asking what specific steps should be taken. The other companies begin planning for the stress test. Some companies’ risk managers actively solicit input from affiliated investment banks and investment advisors concerning the equity markets. Two companies discuss with the product design colleagues what the impact of a dramatic equity decline would be on the behavior of existing annuitants, focusing in particular on whether they would increasingly liquidate their policies. Four companies involve senior management in several of the stress test planning meetings.

192. A decline in the stock market would result in a decrease in the amounts of the variable annuity accounts. If the account values were to fall below the guaranteed amounts under the GMDB contracts, the insurance company would be obligated to make up the difference. *See supra* note 190 and accompanying text.

193. Weber, *supra* note 191, at 61.

Others simply submit the stress test plan to senior management for final approval, and one firm fails to consult senior management at all.

The companies discuss their initial stress test designs with examiners, and explain how their proprietary quantitative statistical models would implement the stress tests. The examiners meet subsequently to de-brief each other on the initial meetings and identify attributes of the stress test, such as consideration of the dynamic interaction of equity market declines and withdrawal of annuitants' funds. They then meet again with the insurance companies and present several of the identified attributes as mandatory best practices for this and future stress tests. Pursuant to the circular, the insurance companies are ordered to review the results of the stress test with senior management and the board of directors. For those insurance companies identified by examiners as having designed unsophisticated stress tests or having failed to consult affiliated expertise or management input, examiners direct their risk management departments to change their stress testing practices in certain respects and notify them that the companies would be subject to spot examinations of stress testing practices with greater frequency.

IV. A BRIEF HISTORY OF THE REGULATION OF FINANCIAL STRESS TESTING

Having presented a normative case in favor of a regulatory regime that fosters deliberation-oriented stress testing, the Article now turns to a brief history of regulatory interventions into stress testing. Have legislators and regulators been aware of the difference between deliberation-oriented stress testing and assurance-oriented stress testing? And to the extent that they have, have their policy initiatives been effective in promoting deliberation-oriented stress testing among regulated financial firms? The following chronology demonstrates that to date, assurance-oriented stress testing has predominated over deliberation-oriented stress testing in guiding regulatory activity.

A. TWO IMPROBABLE EARLY STRESS TESTING INITIATIVES

Regulatory interventions into stress testing practices had improbable origins. The Federal Home Loan Bank Board (FHLBB), which was pilloried for poor performance and eventually disbanded in the aftermath of the savings-and-loan debacle, issued a series of innovative bulletins and policy state-

ments to supervised thrifts concerning their management of interest rate risk and derivatives-related risks in the 1980s.¹⁹⁴ Dramatic spikes in interest rates in the early 1980s led to deposit flight from the thrifts, which were statutorily proscribed from offering competitive interest rates on deposits.¹⁹⁵ Congress eventually liberalized thrift deposit rates and the FLHBB removed the asset restrictions that formerly had prevented thrifts from underwriting non-traditional loan products.¹⁹⁶ But the FHLBB quickly became concerned that thrifts were assuming risks that compromised their safety and soundness as a result of their new investment powers, including by speculating in derivatives.¹⁹⁷ This series of late-1980s FHLBB pronouncements represents the first intervention by bank supervisory authorities into risk management and stress testing regulation. Thrift Bulletin 12, published in 1988, is especially relevant in this context.¹⁹⁸ In it, the agency required not only that thrift boards oversee management's development of a "comprehensive business plan" detailing risk management objectives (including position limits), but also that thrifts perform "sensitivity analysis" before investing in certain instruments.¹⁹⁹

Four years later, another improbable innovator—the U.S. Congress—utilized the concept of stress in its Federal Housing Enterprises Financial Safety and Soundness Act of 1992 (FHEFSSA).²⁰⁰ FHEFSSA established a new Office of Federal Housing Enterprise Oversight (OFHEO) within the Department of Housing and Urban Development as the prudential regulator in charge of monitoring the capital adequacy and safety and soundness of the two major government-sponsored housing entities (GSEs): Federal National Mortgage Associa-

194. LAWRENCE J. WHITE, *THE S&L DEBACLE: PUBLIC POLICY LESSONS FOR BANK AND THRIFT REGULATION* 128–31 (1991).

195. *Id.* at 67–72.

196. The liberalization of the asset restrictions would, the FHLBB hoped, allow thrifts to achieve the required rates of return with higher-yielding assets that they would need to retain fleeing, and attract new, deposits. *Id.* at 72–74.

197. See *Investment Portfolio Policy and Accounting Guidelines*, 54 Fed. Reg. 23,457, 23,461, 23,471 (May 19, 1989).

198. FED. HOME LOAN BANK BD., *THRIFT BULL. NO. 12, MORTGAGE DERIVATIVE PRODUCTS AND MORTGAGE SWAPS* (1988) [hereinafter *THRIFT BULLETIN NO. 12*], available at <http://www.occ.gov/static/news-issuances/ots/bulletins/rescinded-thrift-bulletins/ots-tb-12.pdf>.

199. *Id.* at 2.

200. Housing and Community Development Act of 1992, ch. XIII, Pub. L. 102-550, 106 Stat. 3672, 3941–4012 (codified as amended in scattered sections of 12 U.S.C.).

tion (known as Fannie Mae) and the Federal Home Loan Mortgage Corporation (known as Freddie Mac).²⁰¹ The Act required OFHEO to conduct a simple, periodic scenario test to “measure[] risk in the context of a company’s overall portfolio, including the company’s risk management activities.”²⁰² Congress referred to the scenario as a “stress period.”²⁰³

Under the FHEFSSA stress testing regime, OFHEO would determine the amount of capital each GSE would need to survive a ten-year scenario characterized by large credit losses, large swings in interest rates, and other adverse conditions impacting the assets, liabilities, and off-balance sheet obligations of the GSEs.²⁰⁴ The test would project cash flows from the GSEs’ assets, liabilities, and off-balance sheet obligations and produce monthly balance sheets for the 120 months of the stress test period.²⁰⁵ A GSE would “pass” the stress test, and therefore avoid intrusive interventions by OFHEO, by ensuring a positive total capital position throughout the simulated ten-year adverse period of adverse conditions.²⁰⁶

201. *Id.* §§ 1311–1319G.

202. Organization and Functions, 68 Fed. Reg. 32,627, 32,627 (June 2, 2003); *see also* Press Release, Office of Fed. Housing Enter. Oversight, OFHEO Issues Risk-Based Capital Stress Test Results for Fannie Mae and Freddie Mac 4 (June 27, 2002) [hereinafter OFHEO Stress Testing Press Release], available at <http://www.fhfa.gov/webfiles/2240/62702stresstestresults.pdf> (“OFHEO is among the first financial institution regulators to use its own stress test to determine capital adequacy.”).

203. *See* OFHEO Stress Testing Press Release, *supra* note 202, at 14.

204. Risk-Based Capital, 66 Fed. Reg. 47,730, 47,730–31 (Sept. 13, 2001).

205. Risk-Based Capital, 67 Fed. Reg. 11,850, 11,850 (Mar. 15, 2002).

206. Technically, the GSEs would need to maintain 130% of the total capital required to provide for the extra capital buffer on account of “management and operations risk.” The stress test results were incorporated directly into the machinery of the capital adequacy rules. Under FHEFSSA, the GSEs would be required to maintain core capital of 2.5 percent of assets plus 0.45% of adjusted off-balance sheet obligations *and* total capital in an amount at least equal to the risk-based capital level. Housing and Community Development Act §§ 1361, 1362, & 1364. As used in FHEFSSA, “core capital” consists of common and preferred stocks, paid-in capital, and retained earnings; and “total capital” consists of core capital plus a general (i.e., not applicable to specific provisioned-for losses) foreclosure loss allowance and “any other amounts from sources of funds available to absorb losses incurred by the [GSE]” that the OFHEO director “by regulation determines are appropriate to include in determining total capital.” *Id.* § 1303(4), (18). The “risk-based capital level” (also referred to as “stress test capital” in OFHEO’s implementing rule) for a GSE was the amount of capital that would enable the GSE to survive the 10-year stress test plus 30 percent of that amount to provide for “management and operations risk.” *Id.* § 1361(c); Risk-Based Capital, 66 Fed. Reg. 47,730, 47,731 (Sept. 13, 2001).

It might seem surprising to learn of congressional innovation with avant-garde financial regulatory techniques with respect to Fannie Mae and Freddie Mac. After all, when the housing bubble burst in 2007 nearly three decades after the passage of FHEFSSA, Fannie Mae and Freddie Mac were leveraged sixty-to-one and were placed into conservatorship where they racked up billion-dollar quarterly losses on the taxpayer's dime.²⁰⁷ Closer examination reveals that although the GSE stress tests can be characterized as novel,²⁰⁸ their effectiveness was handicapped by familiar shortcomings of assurance-oriented stress testing methods.

Consider, for example, that: (1) OFHEO, not the GSEs, was responsible for conducting the tests, so no local knowledge was harnessed and GSE corporate governance was unaffected,²⁰⁹ (2) the test was applied with respect to the GSEs' asset portfolios as they existed *at a fixed point in time*,²¹⁰ (3) the variables that were stressed and, in many cases, the methodologies by which they were stressed, were also fixed, specified in FHEFSSA itself,²¹¹ (4) the stress scenarios were drawn from historical precedents, meaning that the stress test assumed, at least with respect to any single variable, that the worst was in the past,²¹²

207. See VIRAL V. ACHARYA ET AL., *GUARANTEED TO FAIL: FANNIE MAE, FREDDIE MAC, AND THE DEBACLE OF MORTGAGE FINANCE* 23–27 (2011).

208. Technically, FHEFSSA was not the trailblazer in establishing regular, periodic stress testing of financial institutions as a component of supervision. In 1991, Congress had already instructed the Farm Credit Administration to establish an Office of Secondary Market Oversight (OSMO) and to direct the OSMO to develop and administer a similar test to the Federal Agricultural Mortgage Corporation (known as Farmer Mac). See Food, Agriculture, Conservation, and Trade Act Amendments of 1991 § 503, Pub. L. No. 102-237, 105 Stat. 1818, 1870–72 (1991) (inserting new section 8.32 in the Farm Credit Act of 1971).

209. OFHEO Stress Testing Press Release, *supra* note 202, at 4.

210. *Id.* at 5.

211. See, e.g., Risk-Based Capital, 66 Fed. Reg. 47,730, 47,732 (Sept. 13, 2001) (noting, for example, that FHEFSSA prescribes credit losses and rises and falls in baseline interest rates to be used in the stress test).

212. OFHEO's treatment of haircuts illustrates this historicity bias. When simulating the stress test's effects on the GSEs' balance sheets, FHEFSSA instructed OFHEO to give the GSEs credit for cash payments that would be received during the stress period from counterparties to GSE contracts, such as mortgage insurance companies and derivatives counterparties. See Risk-Based Capital, 67 Fed. Reg. 11,850, 11,850 (Mar. 15, 2002). However, because the ability of counterparties to pay the GSEs contracted-for cash payments would be adversely affected by the simulated conditions of the stress test, OFHEO provided for percentage-based "haircuts," or discounts, to the value of the cash payments. *Id.* OFHEO determined the rates of counterparty default and the severity levels of the defaults (i.e., the amount of loss a GSE would suffer from

(5) variables were isolated and they did not interact dynamically;²¹³ and (6) the outcome of the test was a binary pass-fail verification that did not prompt further action.²¹⁴ Moreover, by tying the stress test exercise directly to a capital adequacy regime that required compliance with precise rules, Congress entrenched an adversarial, top-down regulatory relationship. OFHEO's responses to the GSEs' comments, published in the Federal Register, reflect a consistently adversarial relationship in which the GSEs, perceiving that increases to capital requirements would reduce their return on equity, advocated at nearly every juncture for the attenuation of the stressed conditions used in the tests.²¹⁵ "Stress" was a periodically negotiated event rather than a continuous subject of deliberation within the firm and between the firm and its regulators.²¹⁶ During these one-off negotiations, the regulator was deliberating on stressed conditions, and the regulated entity minimized threats in order to achieve lower capital requirements.²¹⁷

B. THE DEVELOPMENT OF STRESS TESTING REGULATION FROM 1990-2006

Throughout the 1990s and the first part of the 2000s, stress testing regulation was, with one notable exception, characterized by either assurance-oriented stress testing or a vagueness that defies ready categorization along the deliberation-assurance binary. These regulatory interventions occurred in two principal contexts: regulation pertaining to derivatives risk management and regulation pertaining to capital adequacy.²¹⁸ As with all of bank supervisory policy during this period, the Basel Committee on Banking Supervision—the international bank regulatory policy coordinator—played an important

a default, taking into account, e.g., foreclosure sales of collateral) using data from the Great Depression. *Id.* at 11,851–52.

213. *Id.*

214. Risk-Based Capital, 66 Fed. Reg. 47,730, 47,732 (Sept. 13, 2001). The point is that the capital requirement is either met (i.e., the GSE has sufficient capital in light of the test) or it is not met (i.e., the GSE lacks sufficient capital in light of the test). If the latter, then the GSE must raise additional capital, but there is no follow-up discussion about, for example, what is causing the lack of adequate capitalization or how might those causes be mitigated.

215. Risk-Based Capital, 66 Fed. Reg. 47,730, 47,736 (Sept. 13, 2001).

216. *See, e.g., id.* at 47,742.

217. *See id.* at 47,741–42.

218. BASEL COMM. ON BANKING SUPERVISION, RISK MANAGEMENT GUIDELINES FOR DERIVATIVES ¶¶ III.2, 8, 14 (1994) [hereinafter BASEL RISK MANAGEMENT GUIDELINES], available at <http://www.bis.org/publ/bcbasc211.pdf>.

role.²¹⁹ With the exception of the so-called 1996 market risk amendments to the capital adequacy rules, regulators moved haltingly into this area of regulatory policy, stopping well short of articulating a regulatory vision of deliberation-oriented stress testing. Instead, notwithstanding the promising market risk amendment approach to stress testing, banks during this period used stress tests almost exclusively as an assurance tool, in the process missing key signals of the impending financial crisis.

In 1993, the Office of the Comptroller of the Currency (OCC), which supervises federal chartered national banks, issued Circular No. 277.²²⁰ The circular was noteworthy because with it the OCC made the first attempt at a comprehensive set of risk management guidelines in connection with burgeoning derivatives activities by banks. The OCC directed bank management to, among other things, “facilitate stress testing” in order to “evaluate risk exposures under various scenarios that represent a broad range of potential market movements and corresponding price behaviors and that consider historical and recent market trends.”²²¹ A year later, the Basel Committee published its own guidance on risk management for derivatives activities, mentioning quantitative and qualitative stress tests that “identify[y] possible events or changes in market behavior that could have unfavourable effects on the institution and [assess] the ability of the institution to withstand them.”²²² The Basel Committee reiterated this guidance the following year in a joint report it published with the International Organization of Securities Commissions (IOSCO), instructing bank supervisors to require “[i]nstitutions with significant trading books [to] subject their portfolios on a regular basis to stress tests using various assumptions and scenarios.”²²³ In 1996, the OCC published special guidance concerning credit derivatives and again emphasized the importance of stress testing to “evaluate the

219. *Id.* at Preface ¶ 1; BASEL COMM. ON BANKING SUPERVISION, FRAMEWORK FOR SUPERVISORY INFORMATION ABOUT THE DERIVATIVES ACTIVITIES OF BANKS AND SECURITIES FIRMS 12 (1995) [hereinafter BASEL DERIVATIVES FRAMEWORK], available at <https://www.bis.org/publ/bcbs19.htm>.

220. OFFICE OF THE COMPTROLLER OF THE CURRENCY, BANKING CIRCULAR NO. 277, RISK MANAGEMENT OF FINANCIAL DERIVATIVES (1993).

221. *Id.* at 13.

222. BASEL RISK MANAGEMENT GUIDELINES, *supra* note 218, at ¶III.6.

223. BASEL DERIVATIVES FRAMEWORK, *supra* note 219, at 12.

bank's exposure in a highly stressed market scenario."²²⁴ The following year, in the aftermath of the 1997 Asian crises, the OCC issued supplemental guidance that specifically addressed examiner and board-level oversight of stress testing.²²⁵ In 1998, the Basel Committee and IOSCO updated their 1995 derivatives supervision report by providing some examples of scenarios supervisors might require banks to run.²²⁶

The vagueness of these references to stress testing in the early derivatives risk management guidance contrasts somewhat with the later, more elaborate treatment of stress testing in the Basel Committee-led restructuring of capital adequacy regulation. During the late 1990s and early 2000s, the Basel Committee radically reworked capital adequacy regulation²²⁷ by permitting banks to set their minimum regulatory capital levels by reference to their proprietary statistical models such as VaR.²²⁸ The Basel Committee initially extended such treatment in 1996 to the market risk capital charge—i.e., the amount of capital held against losses on derivatives and other securities that the bank does not expect to hold to term.²²⁹ As a precondition to usage of this new VaR-based market risk capital adequacy regime, banks' risk management systems had to satisfy certain requirements, including having in place a "rigorous and comprehensive stress testing program."²³⁰ The Basel Committee specifically mentioned historical and hypothetical scenario analysis and sensitivity analysis.²³¹ Most interestingly, the

224. OFFICE OF THE COMPTROLLER OF THE CURRENCY, OCC BULLETIN NUMBER 1996-43, GUIDELINES FOR NATIONAL BANKS (1996), available at <http://www.occ.gov/news-issuances/bulletins/1996/bulletin-1996-43.html>.

225. OFFICE OF THE COMPTROLLER OF THE CURRENCY, RISK MANAGEMENT OF FINANCIAL DERIVATIVES (1997), available at <http://www.occ.gov/publications/publications-by-type/comptrollers-handbook/deriv.pdf>.

226. INT'L ORG. OF SECURITIES COMM'NS & BASEL COMM. ON BANKING SUPERVISION, FRAMEWORK FOR SUPERVISORY INFORMATION ABOUT DERIVATIVES AND TRADING ACTIVITIES 12 (1998), available at <http://www.bis.org/publ/bcbs39.pdf>.

227. Capital adequacy rules require banks to maintain minimum amounts of loss-absorbent capital, such as common equity, to serve as a cushion against unanticipated losses. See 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, 789 (2010).

228. See *id.* at 807–11.

229. See BASEL COMM. ON BANKING SUPERVISION, AMENDMENT TO THE CAPITAL ACCORD TO INCORPORATE MARKET RISKS 5–6 (2005), available at <http://www.bis.org/publ/bcbs119.pdf>.

230. *Id.* at 42–43.

231. *Id.*

Committee believed that “stress tests should be both of a quantitative and qualitative nature” and that they should address, in addition to first-order market losses, the “liquidity aspects of market disturbances.”²³² The Committee tied the “qualitative criteria” to deliberation-oriented stress testing themes of (1) use of imagination to identify threats to firm objectives; (2) regular communication of test results to senior management and the board of directors; and (3) dialogue with regulators:

Qualitative criteria should emphasize that two major goals of stress testing are to evaluate the capacity of the bank’s capital to absorb potential large losses and to identify steps the bank can take to reduce its risk and conserve capital. This assessment is integral to setting and evaluating the bank’s management strategy and the results of stress testing should be routinely communicated to senior management and, periodically, to the bank’s board of directors.

....

... Moreover, if the testing reveals particular vulnerability to a given set of circumstances, the national authorities would expect the bank to take prompt steps to manage those risks appropriately²³³

Later that same year, U.S. bank regulators implemented the market risk capital adequacy regime, but they did not specify how the stress tests were to be conducted, leaving it instead to industry to develop best practices for stress testing.²³⁴ Four years later in 2000, the OCC published related guidance on model validation that impliedly touched on stress tests by mentioning the importance of “understanding . . . a model’s strengths and weaknesses.”²³⁵

When the Basel Committee extended this internal models approach to the *credit* and *operational* risk capital charges in 2004, it again reiterated that the statistical models being used to generate required minimum capital levels would need to be subjected to stress testing.²³⁶ With this reform, known as “Basel II” to distinguish it from the initial 1988 Basel Accord on credit risk, the Committee took a step back from deliberation-oriented

232. *Id.*

233. *Id.*

234. See Risk-Based Capital Standards: Market Risk, 61 Fed. Reg. 47,358, 47,363 (Sept. 6, 1996).

235. OFFICE OF THE COMPTROLLER OF THE CURRENCY, OCC BULL. NO. 2000-16, MODEL VALIDATION 2 (2000), available at http://ithandbook.ffiec.gov/media/resources/3676/occ-bl2000-16_risk_model_validation.pdf.

236. BASEL COMM. ON BANKING SUPERVISION, INTERNATIONAL CONVERGENCE OF CAPITAL MEASUREMENT AND CAPITAL STANDARDS 205, 207–08 (2006), available at <http://www.bis.org/publ/bcbs128.pdf>.

stress testing.²³⁷ Basel II required banks to have in place “sound stress testing processes” that identify “unfavourable effects on a [bank]’s credit exposures.”²³⁸ Moreover, the reform imposed a new requirement: banks would have to design and perform a sensitivity analysis—labeled the “credit risk stress test”—subject to regulatory supervision.²³⁹ The Committee’s guidance on the test, however, was hesitant and assurance-focused. For instance, it noted that “the objective is not to require banks to consider worst-case scenarios” and that the test “should . . . consider at least the effect of mild recession scenarios” such as “two consecutive quarters of zero growth to assess the effect on the banks’ [credit exposure].”²⁴⁰ Compared with the Committee’s earlier treatment of market risk stress testing requirements, the Committee’s conservatism in the credit risk context is apparent.

The middle-of-decade period marks an inflection point in the history of regulatory engagement with stress testing. Regulators intermittently touched on stress testing in guidance to supervised institutions. For instance, U.S. bank regulators responded to signs of a bubble in commercial real estate in 2006 by publishing guidance encouraging, but neither requiring nor prescribing specific methodologies for, banks to conduct portfolio-level stress tests so as to “quantify the impact of changing economic scenarios on asset quality, earnings, and capital.”²⁴¹

During this period, regulators discovered that their ability to foster effective stress testing systems within banks had fallen well short of their expectations.²⁴² In 2006, the Board of Governors of the Federal Reserve System (FRB)—the lead U.S. regulator for bank holding company groups—reviewed bank stress testing practices at large, complex banking groups.²⁴³ The FRB conducted the study to facilitate its institutional understanding of the full range of stress testing practices, an urgent task because “there was neither a well-developed set of best practices nor supervisory guidance in this area at the time.”²⁴⁴ The study found that none of the banking groups had an inte-

237. *Id.* at 2–5.

238. *Id.* at 266.

239. *Id.* at 96–97.

240. *Id.*

241. Concentrations in Commercial Real Estate Lending, Sound Risk Management Practices, 71 Fed. Reg. 2302, 2307 (Jan. 13, 2006).

242. GAO *Risk Management Oversight Assessment*, *supra* note 116, at 4.

243. *Id.* at 23.

244. *See id.*

grated stress testing program that incorporated all major financial risks on an enterprise-wide basis.²⁴⁵ Instead, banks were stress testing the impact of adverse events on individual products and business lines rather than on the institution as a whole and missing any dynamic interaction among risk factors or portfolio exposures.²⁴⁶ Even more troubling, none of the groups were regularly conducting worst-case scenario analysis involving insolvency scenarios.²⁴⁷

A Government Accountability Office (GAO) report summarizing the FRB study reads like a tutorial on how an assurance-oriented stress testing approach, and the corresponding de-emphasis of active deliberation, can overwhelm both regulators and risk management departments:

[O]fficials told us that the current crisis had gone beyond what they had contemplated for a worst-case scenario, and they said that they would probably have faced significant resistance had they tried to require the institutions to do stress tests for scenarios such as downgrades in counterparties' credit ratings because such scenarios appeared unlikely. Other regulators raised concerns about stress testing at individual institutions, but we did not find evidence that they had effectively changed the firms' stress testing practices. In the materials we reviewed, one regulator recommended that the institution include worst-case scenarios in its testing. In a 2005 examination report, examiners noted a concern about the level of senior management oversight of risk tolerances. This concern primarily stemmed from lack of documentation, stress testing, and communication of firm risk tolerances and the extent to which these were reflected in stress tests. While the firm later took steps to document formal risk tolerances and communicate this throughout the firm, the recommendation related to stress testing remained open through 2008.²⁴⁸

A 2005 study conducted by a committee at the Bank for International Settlements found that the overwhelming majority of stress tests were being conducted on trading portfolios alone, and that stress testing of credit exposures lagged behind.²⁴⁹ Banks had made little progress at all in stress testing liquidity needs.²⁵⁰ The preponderant use of stress tests remained model validation.²⁵¹ A joint FRB-OCC horizontal review of model validation practices at large banks found that bank practices were deficient even for this basic assurance-oriented stress testing

245. *Id.* at 4.

246. *Id.* at 23.

247. *Id.* at 22–24.

248. *Id.* at 24.

249. BIS, STRESS TESTING PAPER, *supra* note 59, at 9, 14.

250. *Id.* at 14; Taleb et al., *supra* note 165, at 5.

251. BIS, STRESS TESTING PAPER, *supra* note 59, at 5.

function.²⁵² When testifying before Congress in 2009, former Chief National Examiner for the OCC levied a sobering assessment of the extent of managerial deliberation on stress:

As with most other issues, the success of a stress testing program depends importantly on the support and sponsorship provided by senior management. In banks where risk management functions did not perform well, stress testing typically was a mechanical exercise. Management viewed stress tests as more of a “requirement” than an important risk management tool that could lead to internal discussions and debate about whether existing exposures constituted unacceptable risks.²⁵³

These inadequacies in stress testing practices contributed to the financial crisis that started in 2007 and persists today in the form of a debt overhang, particularly in Europe.²⁵⁴

C. A NEW PHASE OF POST-2008 STRESS TESTING REGULATION

Following the recent financial crisis, stress testing regulation has been characterized by two trends that have little in common beyond their invocation of the stress test label.²⁵⁵ On the one hand, regulators have redoubled their isolated, “one-off” regulator-conducted stress tests.²⁵⁶ Regulators initiated this process, and Congress subsequently enshrined it as a mandatory, periodic feature of U.S. bank supervision in the Dodd-Frank Wall Street Reform and Consumer Protection Act of 2010.²⁵⁷ In the European Union, regulators’ use of stress remains more ad hoc, though not less prevalent.²⁵⁸ On the other hand, bank regulators, under the auspices of the Basel Committee, registered their collective failure to embed deliberation-oriented stress testing norms into the corporate governance infrastructure of

252. GAO *Risk Management Oversight Assessment*, *supra* note 116, at 21 (finding banks “lacked requirements for model testing, clearly defined roles and responsibilities for testing, adequate detail for the scope or frequency of validation, and a specific process for correcting problems identified during validation”).

253. Long Statement, *supra* note 45, at 75.

254. See BASEL COMM. ON BANKING SUPERVISION, PRINCIPLES FOR ENHANCING CORPORATE GOVERNANCE 2, 4–5 (2010), available at <http://www.bis.org/publ/bcbs176.pdf>.

255. See *infra* Part IV.C.1–2.

256. See *infra* Part IV.C.1.

257. The Dodd-Frank Wall Street Reform and Consumer Protection Act § 165(i), Pub. L. No. 111-203, 124 Stat. 1376, 1430 (2010).

258. Huw Jones, *BoE’s Cunliffe Says EU Banks Tests Must Be Credible*, REUTERS (March 4, 2014), <http://uk.reuters.com/article/2014/03/04/uk-britain-boe-cunliffe-idUKBREA230S920140304>.

regulated banks and provided more detailed guidance.²⁵⁹ The former trend is solidly an assurance-related effort, and the latter trend is a more ambitious, if undeveloped, deliberation-related effort.

1. Regulator-Conducted Assurance-Oriented Stress Testing

The regulator-conducted stress testing initiatives are a transatlantic phenomenon, and are characterized by several of the hallmarks of assurance-oriented stress testing: partial equilibrium assumptions; static scenarios; precise estimation; and sensitivity to audience perception. In the United States, the early 2009 Supervisory Capital Assessment Program (SCAP) saw U.S. bank regulators conduct a highly publicized, isolated, “one-off” stress test of the nineteen largest bank holding companies. The tests were designed to gauge whether firms would remain sufficiently capitalized in the event of a regulator-designed “worse than expected” recession scenario.²⁶⁰ This adverse scenario tracked real GDP, unemployment and housing prices.²⁶¹ If the results demonstrated to regulators that an additional capital buffer was warranted, the banks would be required to raise capital from private investors or, that failing, the U.S. Treasury.²⁶² Regulators openly admitted that this was not a typical supervisory exercise to gauge banking resilience, and instead was designed in large part in the service of transparency and bolstering confidence.²⁶³ To many in the financial community, the stress tests’ purpose was plainly to reassure the public of the solvency of the sector.²⁶⁴

259. See *infra* Part IV.C.2.

260. See Timothy Geithner, *How We Tested the Big Banks*, N.Y. TIMES, May 7, 2009, at A33.

261. BD. OF GOVERNORS OF THE FED. RESERVE SYS., THE SUPERVISORY CAPITAL ASSESSMENT PROGRAM: DESIGN AND IMPLEMENTATION 6 (2009), available at <http://www.federalreserve.gov/bankinfo/bcreg20090424a1.pdf>.

262. Press Release, Joint Statement by the Treasury, FDIC, OCC, OTS, and the Federal Reserve (Feb. 23, 2009), available at <http://www.federalreserve.gov/newsevents/press/bcreg/20090223a.htm>. The results of the stress tests suggested that the 19 banks would stand to lose about \$600 billion in the event the severe scenario materialized. BD. OF GOVERNORS OF THE FED. RESERVE SYS., THE SUPERVISORY CAPITAL ASSESSMENT PROGRAM: OVERVIEW OF RESULTS 3 (2009), available at <http://www.federalreserve.gov/newsevents/press/bcreg/bcreg20090507a1.pdf>.

263. See, e.g., Geithner, *supra* note 260, at A33.

264. See Andrews & Dash, *supra* note 7, at B4 (quoting a banking analyst’s opinion that bank regulators “are designing [the stress tests] to make it sound like the banking system is in great shape”).

Since 2011, the Dodd-Frank Act has required the FRB to conduct annual stress tests—which form part of the FRB’s new Regulation YY and are referred to as “Dodd-Frank Act Stress Tests” or “DFASTs”—of bank holding companies with total consolidated assets greater than \$50 billion and financial institutions designated by the FSOc as systemically significant.²⁶⁵ Subject companies are referred to as “covered companies.”²⁶⁶ The DFAST program projects revenues, expenses, losses, and the resulting post-stress capital ratios based on three hypothetical scenarios of increasing degrees of adverseness: a baseline scenario, an adverse scenario, and a severely adverse scenario.²⁶⁷ The FRB applies the scenarios in a rigid and uniform manner to all subject companies.²⁶⁸

As noted earlier,²⁶⁹ the Dodd-Frank Act also mandated that covered companies conduct their own DFAST semi-annual stress tests in the first and third quarters (the FRB refers to the third-quarter tests as “mid-cycle” to distinguish them from the first-quarter tests).²⁷⁰ Regulation YY requires covered companies to use the FRB’s own scenarios when conducting annual stress tests.²⁷¹ As such, the company-conducted, first-quarter stress test amounts to little more than an arithmetic verification exercise.²⁷² By contrast, the covered companies themselves are responsible for formulating the scenarios used in the mid-

265. 12 U.S.C. § 5365(a), (i)(1) (2012). The institutions subject to these tests are identical to the set of institutions subject to the internal stress testing requirement. *See supra* note 8.

266. 12 C.F.R. § 252.132(e) (2013).

267. 12 C.F.R. § 252.132 (2013); BD. OF GOVERNORS OF THE FED. RESERVE SYS., DODD-FRANK ACT STRESS TEST 2013: SUPERVISORY STRESS TEST METHODOLOGY AND RESULTS 2 (2013) [hereinafter 2013 DFAST RESULTS], *available at* http://www.federalreserve.gov/newsevents/press/bcreg/dfast_2013_results_20130314.pdf.

268. 2013 DFAST RESULTS, *supra* note 267, at 3.

269. *See supra* note 8 and accompanying text.

270. *See* BD. OF GOVERNORS OF THE FED. RESERVE SYS., DODD-FRANK ACT MID-CYCLE STRESS TESTS 2013: SUMMARY INSTRUCTIONS (2013), *available at* <http://www.federalreserve.gov/newsevents/press/bcreg/bcreg20130513a1.pdf>.

271. *See* 12 C.F.R. § 252.144(b) (2013); Supervisory and Company-Run Stress Test Requirements for Covered Companies, 77 Fed. Reg. 62,378, 62,387 (Oct 12, 2012) (“For the annual stress test, covered companies will use the same scenarios as the Board will use for its supervisory stress analysis.”).

272. It might be objected that the timing of the provision of the results of the annual stress tests makes the requirement more meaningful. Covered companies must provide their results to the FRB by January 5, but the FRB must publish its results only by March 31. *Id.* at 62,383 tbl.1. As a result, the FRB obtains insight into the covered company’s general orientation toward the stress testing exercise.

cycle stress tests.²⁷³ As with the regulator-conducted stress tests, the scenarios must consist, at a minimum, of a baseline scenario, an adverse scenario, and a severely adverse scenario.²⁷⁴ Covered companies are to disclose publicly their pro forma capital levels in the event, however unlikely, that the scenarios come to pass.²⁷⁵

The FRB instructions to firms for the mid-cycle tests are, in marked contrast to the prescriptive, top-down approach to the first-quarter tests, open-ended and indeterminate:

The Board anticipates that covered companies may use a variety of quantitative and qualitative approaches to develop the scenarios. The adverse and severely adverse scenarios used in mid-cycle stress tests should reflect a company's unique vulnerabilities to factors that affect its firm-wide activities and risk exposures, including macroeconomic, market-wide, and firm-specific events. The Board expects the companies to consider their own risk profiles and operations in designing specific elements of the adverse and severely adverse scenarios.²⁷⁶

As with all open-ended regulatory dictates, its effectiveness will depend on the vigor with which the FRB engages with bank management. If these discussions are robust, it should be expected that the FRB-formulated scenarios will be improved. The track record of regulators on this score is not good, though the optimist might see substantial room for improvement.

The results of the DFAST tests are taken into account in a subsequent capital plan review²⁷⁷ process during which the FRB evaluates each covered company's capital plan.²⁷⁸ Specifi-

273. 12 C.F.R. § 252.145(b) (2013).

274. See Supervisory and Company-Run Stress Test Requirements for Covered Companies, 77 Fed. Reg. at 62,387.

275. *Id.* at 62,379. Because this is a forward-looking computation, the rule incorporates assumptions with respect to dividends, repurchases, redemptions, and additional capital raises during the prospective period over which the test spans.

276. *Id.* at 62,387.

277. The capital plan review grows out of, and integrates as a regular feature of banking supervision, the FRB's initial Comprehensive Capital Analysis and Review (CCAR), conducted in 2010. With the CCAR, the FRB assessed the capital adequacy and the internal capital planning processes of the same large, complex bank holding companies that participated in the 2009 SCAP stress test exercise. Policy Statement on the Scenario Design Framework for Stress Testing, 77 Fed. Reg. 70,124, 70,124–25 (Nov. 23, 2012).

278. The FRB requires covered companies to submit annual capital plans for consideration by the FRB. Capital Plans, 76 Fed. Reg. 74,631, 74,633–34 (Dec. 1, 2011). The mandatory elements of a capital plan organize into four components:

(i) an assessment of the expected uses and sources of capital over the planning horizon (at least nine quarters, beginning with the quarter preceding the quarter in which the bank holding company submits its

cally, the FRB assesses whether the covered companies have robust, forward-looking capital planning processes and have sufficient capital to continue operations throughout times of economic and financial stress. If, after consideration of, among other things, the DFAST results, the FRB does not approve of the capital plan, the covered company is prohibited from engaging in any “capital distribution.”²⁷⁹ Under the capital plan review program, the FRB requires that covered companies maintain a minimum 5% core tier 1 capital ratio even under the severely adverse scenario.²⁸⁰ If the covered company cannot demonstrate its ability to meet that standard, the FRB will reject its capital plan.²⁸¹ This power expands the FRB’s formal arsenal beyond its traditional powers under the prompt corrective action regime in place since 1992.²⁸² The capital plan review and the DFAST program represent a substantial strengthening of previous approaches to assessing capital adequacy and promoting best practices at large banks for measuring capital needs and for managing and allocating capital resources.²⁸³

capital plan) that reflects the bank holding company’s size, complexity, risk profile, and scope of operations, assuming both expected and stressful conditions; (ii) a detailed description of the bank holding company’s process for assessing capital adequacy; (iii) the bank holding company’s capital policy; and (iv) a discussion of any expected changes to the bank holding company’s business plan that are likely to have a material impact on the firm’s capital adequacy or liquidity.

Id. at 74,634. The rule requires each covered company capital plan to consider the effects of, in addition to the FRB-formulated stress scenarios, at least one stress scenario developed by the covered company itself. *Id.* at 74,635. The FRB, after review of the capital plan, may approve the plan, disapprove the plan, or require re-submission of the plan. *Id.* at 74,638–41; *see also* Policy Statement on the Scenario Design Framework for Stress Testing, 77 Fed. Reg. at 70,125 (noting that the capital plan review “ties the review of a bank holding company’s performance under stress scenarios to its ability to make capital distributions”).

279. For purposes of its capital plan review rule, the FRB defines “capital distribution” as “a redemption or repurchase of any debt or equity capital instrument, a payment of common or preferred stock dividends, a payment that may be temporarily or permanently suspended by the issuer on any instrument that is eligible for inclusion in the numerator of any minimum regulatory capital ratio, and any similar transaction that the Federal Reserve determines to be in substance a distribution of capital.” *Id.* at 74,637.

280. BD. OF GOVERNORS. OF THE FED. RESERVE SYS., COMPREHENSIVE CAPITAL ANALYSIS AND REVIEW 2013: ASSESSMENT FRAMEWORK AND RESULTS 11 (2013), *available at* <http://www.federalreserve.gov/newsevents/press/bcreg/ccar-2013-results-20130314.pdf>.

281. *See supra* note 278.

282. *See Weber, supra* note 10, at 717 n.287.

283. *Id.* At present, the capital plan review applies to the nineteen largest bank holding companies, while the DFAST program applies to a broader range

In 2010, European Union (EU) authorities established the European Banking Authority (EBA) as an EU-level banking supervisor and charged it with, among other things, administering EU-wide stress tests on EU-domiciled banks.²⁸⁴ The EU lawmaking institutions created the EBA to replace the Committee of European Banking Supervisors (CEBS), which had been created in 2004 to encourage supervisory coordination and information sharing among EU member state bank supervisors and promote consistent implementation of EU banking laws and regulation.²⁸⁵ As problems in peripheral EU state banking sectors attracted attention, CEBS conducted an initial round of stress tests in May of 2009 that modeled gross domestic product, unemployment rates, and real estate market prices.²⁸⁶ The CEBS conducted a second round of tests in 2010 and the newly formed EBA conducted its first round of tests in 2011.²⁸⁷ The EU stress tests operate similarly to the FRB's capital plan review program; banks were required in July 2011 to raise additional core tier 1 capital to the extent that they would have less than 5% core tier 1 capital ratio under the adverse scenario.²⁸⁸ In December of 2011, the EBA raised its required threshold to

of institutions based on the statutory threshold. Press Release, Bd. of Governors of the Fed. Reserve Sys., Stress Tests and Capital Planning, (Feb. 25, 2014), available at <http://www.federalreserve.gov/bankinforeg/stress-tests-capital-planning.htm>.

284. See Council Regulation 1093/2010, art. 20(2)(b), 2010 O.J. (L 331) 12, 29 (EU). Specifically, the EBA stress tests are intended "to assess the resilience of financial institutions . . . to adverse market developments, and [and to] evaluate the potential for systemic risk to increase in situations of stress, ensuring that a consistent methodology is applied at the national level to such tests." *Id.* As of mid-2013, there is speculation that the EBA will be dissolved in connection with the creation of a Eurozone-wide banking union. Patrick Jenkins, *European Banks Lift Capital Buffers*, FIN. TIMES, July 12, 2012, at 16.

285. *About CEBS, CEBS Archive*, EUROPEAN BANKING AUTHORITY, <http://www.eba.europa.eu/cebs-archive> (last visited Apr. 21, 2014).

286. Press Release, Comm. of Eur. Banking Supervisors, Results of the EU-Wide Stress Testing Exercise, (Oct. 1, 2009) available at <http://www.eba.europa.eu/documents/10180/15977/CEBS-2009-180-Annex-2-%28Press-release-from-CEBS%29.PDF>.

287. EUR. BANKING AUTH., 2011 EU-WIDE STRESS TESTS: AGGREGATE RESULTS 3 (2011) [hereinafter AGGREGATE RESULTS], available at http://www.eba.europa.eu/documents/10180/15935/EBA_ST_2011_Summary_Report_v6.pdf/54a9ec8e-3a44-449f-9a5f-e820cc2c2f0a.

288. *Id.* at 38.

9%, making it decidedly more stringent than the U.S. capital plan review.²⁸⁹

These stress tests illustrate several shortcomings with assurance-oriented stress testing. The first shortcoming is a danger against which all stress tests should guard: undue optimism. For example, the dire scenario regulators presented as a “more adverse scenario” in the February 2009 SCAP program appeared anodyne by April of 2009.²⁹⁰ The 2010 CEBS stress test exercise in Europe gave a clean bill of health to all but seven of the ninety-one tested banks, identifying an aggregate capital shortfall of only €3.5 billion.²⁹¹ Four months later, the European Central Bank, the International Monetary Fund, and the European Union announced a bailout of Ireland that allocated €10 billion—and made available another €25 billion—to recapitalize the banks of Ireland alone!²⁹² Furthermore, Greek banks and Spanish *cajas* (savings banks) passed the 2010 CEBS stress tests, just months before requiring unprecedented levels of public assistance.²⁹³ Banks from the most recent Eurozone bailout recipient, Cyprus, passed the stress tests in 2010 and 2011 because the adverse scenario did not contemplate the possibility of losses on government bonds.²⁹⁴ This was no minor omission; the primary cause of uncertainty in the banking sector at the time was exposure to sovereign bonds.²⁹⁵ The Franco-

289. Patrick Jenkins, *Pretending the World Is Fixed Is a Gamble*, FIN. TIMES, Mar. 26, 2012, <http://www.ft.com/intl/cms/s/0/7cc387cc-775c-11e1-93cb-00144feab49a.html>.

290. See Wessel, *supra* note 2, at A2.

291. Patrick Jenkins & Brooke Masters, *Bank Watchdog Sets out to Square the Circle*, FIN. TIMES, Feb. 14, 2011, <http://www.ft.com/intl/cms/s/0/d70dd886-3865-11e0-959c-00144feabdc0.html>.

292. See Patrick Jenkins, *Bail-out Document Runs Dry on Tackling Liquidity Issue*, FIN. TIMES, Nov. 30, 2010, at 20.

293. See David Enrich, *New Doubts on EU Bank Stress Tests*, WALL ST. J., July 20, 2010, <http://online.wsj.com/news/articles/SB10001424052748704720004575377202517842246> (noting that stress tests by market researchers at another bank estimated that the *cajas* alone would face a €50 billion shortfall in the event of a severe downturn).

294. David Enrich & Charles Forelle, *Greek Bets Sank Top Lenders*, WALL ST. J., Mar. 27, 2013, <http://online.wsj.com/news/articles/SB10001424127887323501004578386762342123182>.

295. Patrick Jenkins, *Power to the Regulators*, FIN. TIMES, Dec. 12, 2011, <http://www.ft.com/intl/cms/s/0/c1266a8a-24e4-11e1-8bf9-00144feabdc0.html> (“When the EBA ran [its 2011] stress test[s] . . . it applied bleak forecasts for what might happen to European economic growth, unemployment and property prices, but in the face of political hostility from European Union leaders, it stopped short of applying market valuations of peripheral sovereign debt—

Belgian bank Dexia passed the 2011 stress test, only to require France and Belgium to assure investors of their readiness to guarantee Dexia's obligations three months later.²⁹⁶ Eventually, in December of 2011, the EBA updated its stress tests to take into account a modest loss on sovereign bond holdings.²⁹⁷ Excessive optimism is not limited to scenario design. The U.S. capital plan review program's requirement that covered companies maintain tier 1 capital ratios of 5%, after all, sets a bar that is actually lower than the Basel III capital ratios that U.S. bank regulators agreed to implement by 2015.²⁹⁸

A second shortcoming, which is related to the first, flows from having hybrid intended audiences: on the one hand, the tests were designed to advert supervisors to problem institutions. But on the other hand, they were designed to communicate to markets and investors that the financial system was sound.²⁹⁹ Several industry commentators had expressed skepticism about the "adversity" of the "more adverse scenario" on these grounds.³⁰⁰ A Royal Bank of Scotland economist voiced the concern, stating "If you do have the Greek banks pass the tests, there may still be some skepticism in the markets about whether or not they've been rigorous and whether they've been a true test of the system."³⁰¹ Arguing in favor of "realistic" severe scenarios, *Financial Times* banking commentator Wolfgang Münchau discusses the tension between the political and diagnostic functions of the EBA stress tests:

precisely the issue that was at the heart of investors nervousness about banks' financial strength.").

296. Jill Treanor, *How Did Europe's Bank Stress Tests Give Dexia a Clean Bill of Health?*, THE GUARDIAN BUSINESS BLOG (Oct. 5, 2011, 4:10 PM), <http://www.guardian.co.uk/business/blog/2011/oct/05/europe-bank-stress-tests-dexia>.

297. Jenkins, *supra* note 295, at 14.

298. Regulatory Capital Rules: Regulatory Capital, Implementation of Basel III, Minimum Regulatory Capital Ratios, Capital Adequacy, Transition Provisions, and Prompt Corrective Action, 77 Fed. Reg. 52,792, 52,824 tbl.9 (Aug. 30, 2012).

299. Jenkins & Masters, *supra* note 1 ("The EBA's aims are twofold: first, to strengthen the system by pushing banks that are thinly capitalised relative to their underlying risks into raising fresh equity: and second, to convince the world about that strength—helping investors, particularly from outside Europe, to differentiate among eurozone banks and stop shunning them en masse.").

300. See Andrews & Dash, *supra* note 7, at B1, B4 (reporting that a bank analyst found the more adverse scenario assumptions "too optimistic" and "[not] harsh enough"—in particular with respect to the expected growth to be generated from 2009 federal stimulus spending).

301. Enrich, *supra* note 293.

First, by realistic stress I mean the inclusion of extreme, not probable, worst-case scenarios. Given the recent discussions about Greece, this must include the worst estimates of a “haircut”—a deduction suffered by bondholders—of about 50 per cent of the face value of Greek bonds. The stress tests will, according to reports last week, include a uniform haircut on sovereign bonds of 3 per cent. This number is a joke. Some institutions will have a stronger exposure to Greece, Portugal, Ireland or Spain than others, and it is important that those banks are stressed on the assumption of significant haircuts of their sovereign risk portfolios. I can see the politics behind the 3 per cent figure. It is official EU policy to deny the reality that Greece might default or restructure. A genuine stress test might expose the EU’s position as indefensible. Those opposed to any inclusion of sovereign risk into the stress tests argue that the mere assumption of a haircut might turn into a self-fulfilling prophecy. The market would jump to the wrong conclusions. This is a silly argument. Stress test scenarios are not forecasts. They are only scenarios, of the kind that market participants have already factored into the pricing of bonds.³⁰²

A third shortcoming is that the scenarios are drawn from historical precedents. For example, in 2011 the EBA used loan loss provision and profitability data from 2009 as a benchmark in setting the “adverse” scenario.³⁰³ Historical scenarios are not intrinsically suspect, and their use is a key part of any stress testing program; to eliminate them would be to avoid learning from history. Nevertheless, the regulatory imposition of a unitary scenario on all banks (as occurs under both the EBA and DFAST test programs) means that banks and regulators are not engaged in a dialogic, iterative process of *scenario development* that can harness the imagination of multiple actors. The need to develop a scenario leads naturally to an overreliance on historical precedents, which are identifiable, numerous, and as to which regulators have reliable data.

A fourth shortcoming is that these stress tests are static and assume partial equilibrium models of the financial system. As a result, they do not capture vulnerabilities to liquidity crunches and other more dynamic phenomena. The FRB acknowledges that the DFAST tests are limited by their partial equilibrium assumptions: “These projections incorporate a number of conservative modeling assumptions, but do not make explicit behavioral assumptions about the possible actions of a BHC’s creditors and counterparties in the scenario, except through the severely adverse scenario’s characterizations of fi-

302. Wolfgang Münchau, *Europe Risks Failing the Real Test on Banks*, FIN. TIMES, Jul. 5, 2010, <http://www.ft.com/intl/cms/s/0/a989f262-879b-11df-9f37-00144feabdc0.html>.

303. AGGREGATE RESULTS, *supra* note 287.

financial asset prices and economic activity.”³⁰⁴ These assurance-related shortcomings limit the effectiveness of these otherwise encouraging programs.

2. Basel Committee Principles: An Overture to Deliberation-Oriented Stress Testing?

While the global financial system was still in the throes of the 2008 financial crisis, the Basel Committee published a series of stress testing “principles” to guide bank management and bank supervisors in their efforts to promote effective stress testing programs.³⁰⁵ Noting that it had already required banks to maintain rigorous stress testing programs in place, the Committee registered its disappointment with the track record of stress testing.³⁰⁶ The document’s definition of stress test is noteworthy: “A stress test is commonly described as the evaluation of a bank’s financial position under a severe but plausible scenario to assist in decision making within the bank.”³⁰⁷ As discussed at length in Part IV.B, it is, in light of the predominance of assurance-oriented stress testing regulatory efforts, an exaggeration to describe the connection between stress testing and firm decision-making as a “commonly” shared precept. Nevertheless, in the document the Committee envisages a new role for bank regulators that resonates with deliberation-oriented stress testing. The supervisory principles are set forth below, with brief descriptions of how they might be used to foster deliberation:

- (1) **Supervisors should make regular and comprehensive assessments of a bank’s stress testing program.**³⁰⁸ By regularly reviewing stress test programs and scenarios, regulators will be able (1) to identify those “good actor” institutions that deliberate more actively on the possibilities of failure and (2) to present scenarios identified by the good actors prospectively to institutions with assurance-oriented stress testing mindsets.
- (2) **Supervisors should require management to take corrective action if material deficiencies in the stress testing program are identified or if the results of stress tests are not adequate-**

304. 2013 DFAST RESULTS, *supra* note 267.

305. BCBS, STRESS TESTING PRINCIPLES, *supra* note 46.

306. *Id.* at 1 (lamenting the failure of bank stress tests to “produce large loss numbers in relation to their capital buffers . . . or their actual loss experience” and to “include[] more severe scenarios”); *see also supra* notes 230–41 and accompanying text (discussing the Committee’s actions to require rigorous stress testing programs).

307. BCBS, STRESS TESTING PRINCIPLES, *supra* note 46, at 2.

308. *Id.* at 17.

ly taken into consideration in the decision-making process.³⁰⁹

Here, the Committee expressly links the stress testing program to decision-making and corporate governance. If stress testing is limited to a verification process performed by risk management departments without interaction with other business units, regulators should be empowered to intervene.

(3) **Supervisors should assess and, if necessary, challenge the scope and severity of firm-wide scenarios, including by requiring reverse stress tests.** The reference to reverse stress tests—which intrinsically require active deliberation on the channels through which failure might emerge³¹⁰—again suggests a deliberation-oriented stress testing approach.

(4) **Supervisors should examine banks' stress test results when assessing whether banks have adequate capital and liquidity on a forward-looking basis.** The Committee notes that “[s]upervisors should . . . be able to understand the rationale for management decisions to take or not to take remedial actions” in response to the results of its stress tests. This again implies regular dialogue with bank management concerning stress and failure and responses thereto. With respect to liquidity risk management, which is, as discussed above, rarely susceptible to assurance-type modeling, the Committee notes that regulators “should review the use of stress test results to ensure that the potential impact on a bank’s liquidity is fully considered and discussed at senior management level.”³¹¹

(5) **Supervisors should be adequately resourced and technically able, and should engage in dialogue with other public authorities and industry to identify systemic vulnerabilities and the ways in which they might unfold.**³¹² The Committee states its hope that regulators and industry engage in a “constructive, systematic dialogue” concerning “which scenarios could unfold and [which] systemic interactions could crystallise.”³¹³ If regulators have the ability and resources to contribute meaningfully to this dialogue, including by resisting tendencies to oversimplify, they will contribute to a better understanding³¹⁴ of the causal environment in which financial institutions operate.

309. *Id.*

310. *See supra* Part II.C.1.b.

311. BCBS, STRESS TESTING PRINCIPLES, *supra* note 46, at 19.

312. *Id.*

313. *Id.*

314. Each of these supervisory principles is explained in greater detail in the Basel Committee document. *See id.* at 17–19. A sixth supervisory principle instructs regulators to “consider implementing stress test exercises based on common [regulator-provided] scenarios” so as to complement banks’ own internal stress tests and better enable regulators to understand the possible impact of specific stress events. *Id.* at 19. The EBA stress tests and the FRB post-crisis activities respond directly to this charge, *see supra* notes 264–68, 284 and accompanying text, but the Basel Committee emphasizes that these “supervisory stress tests should on their own not be considered as sufficient” and that regulators “should make clear that these are not a substitute for stress tests designed by bank management, given that a common supervisory scenar-

These principles espouse expectations that bank regulators will learn from the financial crisis and adopt a deliberation-oriented stress testing approach. The OCC's Chief Bank Examiner has publicly touted the OCC's role in developing these Basel principles.³¹⁵ And the Basel Committee has reiterated them in subsequent publications.³¹⁶ The EBA stress tests in Europe and the DFAST and capital plan review program in the United States demonstrate, however, that regulators have a long way to go to operationalize the principles.

V. THREE PRINCIPLES TO GUIDE THE USE OF REGULATION TO INSTITUTIONALIZE DELIBERATION- ORIENTED STRESS TESTING

The limited progress to institutionalize deliberation-oriented stress testing—both at the regulatory level and at the industry level—is to some extent understandable. Deliberation on remote possibilities of future contingent states, however catastrophic, does not come naturally.³¹⁷ Sociologist Carol Heimer has coined the term “bounded imagination” to describe how “people . . . have rather limited capacities to imagine alternatives to the one that exists or that they have chosen.”³¹⁸ The problem is only exacerbated in bureaucratic environments with multiple corporate actors, stakeholders, and goals—or where, as with large global corporations, managers face competing demands on their attention spans in a high-pressure environment.

The report on the 1986 space shuttle Challenger disaster notes that when it came to formulating pre-launch estimates of

io [such as are provided in the DFAST exercise] is not tailored to the unique characteristics of individual banks.” BCBS, STRESS TESTING PRINCIPLES, *supra* note 46, at 19.

315. See Long Statement, *supra* note 45, at 79.

316. See, e.g., BASEL COMM. ON BANKING SUPERVISION, ENHANCEMENTS TO THE BASEL II FRAMEWORK 25 (2009), available at <https://www.bis.org/publ/bcbs157.pdf> (“Supervisors should assess the effectiveness of banks’ stress testing programme in identifying relevant vulnerabilities. Supervisors should review the key assumptions driving stress testing results and challenge their continuing relevance in view of existing and potentially changing market conditions. Supervisors should challenge banks on how stress testing is used and the way it affects decision-making. Where this assessment reveals material shortcomings, supervisors should require a bank to detail a plan of corrective action.”).

317. See PARKER, *supra* note 10, at 60–61.

318. Carol A. Heimer, *Legislating Responsibility* 25 (Am. Bar Found., Working Paper No. 9711, 1997).

the probability of a loss of vehicle and of human life, the shuttle engineers and the NASA management had wildly different intuitions.³¹⁹ The engineers, true to their safety-minded professional ethos, estimated odds of mission failure up to 1000 times higher than those of managers.³²⁰ After discussing this discrepancy, the report poses the obvious question: “What is the cause of management’s fantastic faith in the machinery?”³²¹ Financial regulators could pose a similar question with respect to managers at the firms they supervise: what is the cause of management’s fantastic faith in their risk management systems?³²² More important for purposes of this Article, however, is the likely follow-up question: “How can we use regulations in a manner that encourages managers, like engineers, to deliberate more actively on worst-case scenarios and, more generally, failure?” This Part takes up that question.

It goes without saying that using public regulatory power to institutionalize deliberation-oriented stress testing within financial firms presents unique challenges. The open-ended and indeterminate character of its outputs is a poor fit for the traditional administrative agency’s toolkit.³²³ To combat this dilemma, regulators should bear in mind three themes in their dialogue with regulated firms over stress testing. First, they should understand their task as involving management-based regulation—a regulatory approach that acts on corporate *planning processes* rather than demanding specific technologies or outputs. Second, they should encourage the further development of the “quantitative skepticism” that Anette Mikes, in her field research with bank risk management departments, has noted already exists in many large financial institutions.³²⁴ Deliberation-oriented stress testing will have an easier time gaining traction in a quantitatively skeptical risk management department.³²⁵ Third, regulators should draw from the examples provided by HROs such as nuclear power plants, aircraft carri-

319. II PRESIDENTIAL COMM’N ON THE SPACE SHUTTLE CHALLENGER ACCIDENT, REPORT TO THE PRESIDENT app. F, at F-1 (1986).

320. *See id.*

321. *Id.*

322. *See* BCBS, STRESS TESTING PRINCIPLES, *supra* note 46, at 2 (“Stress testing practices at most banks . . . did not foster internal debate nor challenge prior assumptions such as the cost, risk and speed with which new capital could be raised or that positions could be hedged or sold.”).

323. *See* Weber, *supra* note 59, at 7.

324. *See supra* notes 56–58 and accompanying text.

325. *See supra* notes 57–58 and accompanying text.

ers, air traffic control systems, and wildfire-fighting units. Management theorists have admired the ability of these HROs to maintain reliable, resilient performance even in conditions of uncertainty and volatility.³²⁶ In other words, the success of stress testing as a regulatory tool depends on whether regulated institutions implement HRO decision-making norms.³²⁷

A. MANAGEMENT-BASED REGULATION: UNDERSTANDING WHY STRESS TESTING MAKES SENSE AS A REGULATORY MECHANISM

The wellness of fit between a regulatory instrument and a regulatory objective is a question that often goes under-examined in the rush to construct a regulatory solution to a perceived market failure.³²⁸ Once the regulatory objective is determined, however, the choice of instrument is the most critical task.³²⁹ While the appeal of stress and failure deliberation is intuitive, reflection on regulatory design—thinking about the way we think about a regulatory tool—is indispensable.

Cary Coglianese and David Lazer present a typology of regulatory mechanisms that is instructive in this context. They describe a traditional dichotomy between technology-based regulation (TBR) and performance-based regulation (PBR) that dates to now-Justice Stephen Breyer's academic administrative law work.³³⁰ When administering TBR tools, regulators specify techniques, procedures, and restrictions.³³¹ When administering PBR tools, regulators intervene at the output stage,³³² as with Pigouvian taxes.³³³ TBR tools are useful where the regulated entities are relatively homogeneous and regulatory outputs—i.e., the measure of social harm or good that prompted the regulatory intervention in the first place—are difficult to monitor.³³⁴ By contrast, PBR tools work best where circumstances are het-

326. See *supra* note 14.

327. See *supra* note 15.

328. Cf. Robert Baldwin & Julia Black, *Really Responsive Regulation*, 71 MOD. L. REV. 59, 59 (2008) (discussing the difficulties inherent in regulatory response).

329. *Id.* (emphasizing that consideration of a regulatory approach must begin by asking “whether it offers assistance in addressing the challenges regulators face in practice”).

330. See Coglianese & Lazer, *supra* note 12, at 691–92.

331. See *id.* at 694.

332. See *id.*

333. Cf. PAUL KRUGMAN ET AL., *ESSENTIALS OF ECONOMICS* 271–73 (2d ed. 2010) (describing Pigouvian taxes).

334. See Coglianese & Lazer, *supra* note 12, at 705.

erogeneous because they avoid frictions involved when TBR tools impose uniform, one-size-fits-all technologies.³³⁵ That said, in order for PBR tools to function the regulator must be able to monitor outputs in order to calibrate regulatory response in a valid manner; without easily monitored outputs, PBR tools will not work.³³⁶

To Coglianesi and Lazer, the TBR-PBR dichotomy therefore provides no guidance to policymakers where the market is at once characterized by heterogeneous market actors and difficult-to-monitor regulatory outputs.³³⁷ Their solution is to propose management-based regulation (MBR) as a third model of regulatory mechanism design.³³⁸ MBR describes when regulators direct market actors to engage in a planning process that aims to achieve public goals.³³⁹ Market actors enjoy the freedom to fashion solutions to problems in ways that are sensitive to their local circumstances.³⁴⁰ The regulatory intervention occurs at the planning stage, with regulators helping and overseeing firms as they deliberate on how to best promote regulatory objectives.³⁴¹ It offers the decentralized context-specificity of PBR without relying on precise measurements of outputs.³⁴² Where policymakers have largely abandoned structural regulation and the regulated market is subject to endemic complexity, MBR is a naturally attractive approach.³⁴³

The idea of MBR resonates with many other theoretical orientations to administrative law and practice, including—to name a few—meta risk management,³⁴⁴ new governance,³⁴⁵ ex-

335. *See id.* at 725 (emphasizing the “flexibility” or performance-based regulation).

336. *See id.*

337. *See id.* at 705.

338. *See id.* at 725–26.

339. *See id.* at 694.

340. *Cf. id.* at 726 (“Management-based regulation may be the best available regulatory approach for problems that require fine-grained analysis of local circumstances.”).

341. *See id.* at 694, 706.

342. *See id.* at 692.

343. *Cf. id.* at 702 (noting that MBR can be particularly important “especially with respect to problems that arise from breakdowns in complex systems or that require coordination among a large number of interactive human and technological processes”).

344. *Cf. John Braithwaite, Meta Risk Management and Responsive Regulation for Tax System Integrity*, 25 *LAW & POLY* 1, 1–2 (2003) (introducing meta risk management as the concept of assessing or regulating the risk management practices of corporations).

perimental governance,³⁴⁶ enforced self-regulation,³⁴⁷ contextualizing regimes,³⁴⁸ collaborative governance,³⁴⁹ and meta-regulation.³⁵⁰ It is not merely the TBR/PBR dichotomy that is at risk; the administrative state as traditionally conceived is contested, as governance is pushed down and throughout what used to be termed the private sector. Rather than mandating specific procedures or outcomes, public power is used to influence the “attitudinal settings” of regulated firms that determine how they process, produce, deliberate on, and react to information.³⁵¹ Policymakers seek to regulate the exercise of judgment—in the context of risk management, they engage in meta-risk management, or the risk management of risk management.³⁵²

345. Cf. Orly Lobel, *The Renew Deal: The Fall of Regulation and the Rise of Governance in Contemporary Legal Thought*, 89 MINN. L. REV. 342, 343–44 (2004) (discussing the “paradigm shift from a regulatory to a governance model”); David M. Trubek & Louise G. Trubek, *New Governance & Legal Regulation: Complementarity, Rivalry, and Transformation*, 13 COLUM. J. EUR. L. 539, 541 (2006) (“These processes, which we will collectively label ‘new governance’, may encourage experimentation; employ stakeholder participation to devise solutions; rely on broad framework agreements, flexible norms and revisable standards; and use benchmarks, indicators and peer review to ensure accountability.”).

346. Cf. Michael C. Dorf & Charles F. Sabel, *A Constitution of Democratic Experimentalism*, 98 COLUM. L. REV. 267, 288 (1998) (discussing experimental governance at local levels); Charles F. Sabel & William H. Simon, *Minimalism and Experimentalism in the Administrative State*, 100 GEO. L.J. 53, 55 (2011) (“In experimentalist regimes, central institutions give autonomy to local ones to pursue generally declared goals.”).

347. Cf. IAN AYRES & JOHN BRAITHWAITE, RESPONSIVE REGULATION 102–03 (1992) (discussing enforced self-regulation as “negotiations between state and *individual* firms that result in flexible, particularistic standards and enforcement”).

348. Cf. Charles F. Sabel & William H. Simon, *Contextualizing Regimes: Institutionalization as a Response to the Limits of Interpretation and Policy Engineering*, 110 MICH. L. REV. 1265, 1266 (2012) (“[I]nstead of making the decision directly, the officials charged with decisionmaking adopt the normative output of one or more specialized bodies of stakeholders.”).

349. Cf. Jody Freeman, *Collaborative Governance in the Administrative State*, 45 UCLA L. REV. 1, 22 (1997) (discussing general features of collaborative governance, such as flexibility and interdependency).

350. Cf. PARKER, *supra* note 10, at 245–91 (discussing the regulation of corporate decision making).

351. See Baldwin & Black, *supra* note 328, at 68–70; David Hess & Cristie L. Ford, *Corporate Corruption and Reform Undertakings: A New Approach to an Old Problem*, 41 CORNELL INT’L L.J. 307, 327–29 (2008).

352. See Kenneth A. Bamberger, *Technologies of Compliance: Risk and Regulation in a Digital Age*, 88 TEX. L. REV. 669, 681 (2010); Braithwaite, *supra* note 344, at 2–3.

But what does it mean to re-focus the regulatory lens on the planning processes instead of specific technologies (as with TBR) or mandated outcomes (as with PBR)? On the one hand, this new focus on the planning stage could take the form of a one-step verification of corporate processes. On the other hand, it could focus on continuous learning, benchmarking, and identifying best practices on a rolling basis.³⁵³ This latter model of MBR regulation draws from “experimentalist” theories of governance that are influenced by the pragmatist precept of the reciprocal determination of ends and means.³⁵⁴ The ends of financial stability and safety and soundness are impacted by the means by which risk managers and executives implement corporate risk governance systems. As such, solutions to regulatory problems must be seen as provisional rather than static. Regulatory action should be “conversational” in nature and responsive not only to changing markets but also to the institutional and functional contexts within which regulation occurs.³⁵⁵ To a large degree, the choice between viewing MBR as process verification or as continuous learning represents the choice between assurance-oriented stress testing and deliberation-oriented stress testing. And, for the same reasons applicable in that context, continuous learning is preferable to periodic verification.³⁵⁶

Another concern to bear in mind is that opening up regulation to participation by regulated firms presents the risk of a legitimacy deficit.³⁵⁷ Whereas traditional administrative law provides for ample private sector participation, such participation is channeled through a pluralistic notice-and-comment process.³⁵⁸ By contrast, an experimentalist MBR approach to regulation is both too decentralized and too provisional to permit a formal notice-and-comment process involving all affected stakeholders, including users of credit, providers of credit (particularly depositors and money market investors), and even

353. William H. Simon compares this sort of legal-regulatory regime to Toyota’s highly flexible, nonhierarchical production system—and in particular its commitment to having zero tolerance to failure and its ability to adapt to changing circumstances. See William H. Simon, *Toyota Jurisprudence: Legal Theory and Rolling Rule Regimes*, in *LAW AND NEW GOVERNANCE IN THE EU AND THE US* 37, 44–45 (Gráinne de Burca & Joanne Scott eds., 2006).

354. See Dorf & Sabel, *supra* note 346, at 284–87.

355. See JULIA BLACK, *RULES AND REGULATORS* 37–44 (1997).

356. See *supra* Part III.B.

357. See Weber, *supra* note 227, at 850–52.

358. See Freeman, *supra* note 349, at 11–12.

taxpayers in their capacity as the ultimate providers of government guarantees.³⁵⁹ Instead, a better prototype is the frequent and informal regulator-regulatee interaction of the bank examination. Because the bank examination occurs outside the gaze of the public, care must be given to ensure that MBR initiatives such as deliberation-oriented stress testing are not captured by industry on account of its privileged access to regulators and role as information-provider.³⁶⁰

Direct stakeholder participation may very well prove unworkable in this context. Nevertheless, although MBR necessarily entails extensive involvement of industry actors in planning regulation, the legitimacy deficit is not a structural, endemic problem. It is at least theoretically possible for an internal industrial morality—which Neil Gunningham and Joseph Rees define as “an industry-wide normative framework, a set of industrial principles and practices that defines right conduct as it spells out the industry’s public commitment to moral restraint and aspiration”³⁶¹—to attenuate these concerns. MBR is therefore not only about innovating the regulatory toolkit by changing the *instruments of regulators*; it also relies to some extent on altering the *objectives of regulated firms*. Philip Selznick makes this point:

If an organization has a well-developed internal morality—driven by the quest for excellence, sustained by the interplay of means and ends—the community’s strategy may well shift from *external to internal control*. Instead of demanding conformity to standards imposed by legislation and regulation, we may place greater reliance on moral development. In this way, the internal morality of an institution becomes a resource for public policy.³⁶²

The introduction of morality into the discussion both complicates and facilitates MBR regulatory programs. It facilitates a program to the extent that it offers a way to fill the legitimacy gap. But it complicates a program to the extent it delimits

359. *Cf. id.* at 83 (“[P]ursuit of collaboration requires efforts to transcend the familiar debates over controlling agency discretion and depends upon a willingness to experiment with nontraditional forms of accountability.”).

360. *See Weber, supra* note 227, at 850–54.

361. Neil Gunningham & Joseph Rees, *Industry Self-Regulation: An Institutional Perspective*, 19 *LAW & POLY* 363, 376 (1997); *see also* PARKER, *supra* note 10, at 295–97 (stating that harnessing industrial morality for public regulatory objectives depends in part on good relations with corporate stakeholders, bolstered by an “open system” dialogue with them).

362. PHILIP SELZNICK, *THE MORAL COMMONWEALTH* 327 (1992). Selznick’s use of “internal morality” is similar to Gunningham and Rees’s use of “industrial morality.” *Compare id.*, with Gunningham & Rees, *supra* note 361, at 376.

the range of possible subjects for which the MBR regulatory program will work. An MBR approach to stress testing that seeks to foster increased mindfulness of stress and failure will only achieve regulatory objectives for those institutions that are willing to deliberate meaningfully on stress and failure. A precondition to an effective MBR regulatory initiative is therefore the development of an industrial morality, a process calling to mind to what Christine Parker refers to as the “institutionalization of responsibility.”³⁶³ The next section takes up the question of what an appropriate industrial morality for deliberation-oriented stress testing might look like.

B. AN INDUSTRIAL MORALITY OF QUANTITATIVE SKEPTICISM

The regulation of deliberation-oriented stress testing is, at bottom, about influencing the moral character of financial institutions so that they maintain capital at levels sufficient to ensure resilient performance during periods of stress.³⁶⁴ Such an orientation recognizes that risk management techniques will only work if attention is paid to the organizational environment within which they operate.³⁶⁵ But what does it mean to speak of an institution’s character? Selznick’s discussion of institutional character is instructive:

As applied to institutions, “character” is a broader idea than “culture.” Culture is the symbolic expression of shared perception, valuation, and belief. Therefore the idea of “organizational culture” properly emphasizes the creation of common understandings regarding purpose and policy. The character of an organization includes its culture, but something more as well. A pattern of dependency—for example, on a specific labor force, a market, or particular suppliers—

363. See PARKER, *supra* note 10, at 61 (“The success of corporate regulation depends crucially on a corporation’s ability to institutionalize responsibility.”).

364. See GRP. OF THIRTY, FINANCIAL REFORM: A FRAMEWORK FOR FINANCIAL STABILITY 43 (2009) (“Regulators will need to encourage banks to internalize [the] discipline [of maintaining adequate capitalization during expansion periods] by requiring capital management policies to be tied to careful analysis of what stress scenarios imply about capital needs.”).

365. See C.F. LARRY HEIMANN, ACCEPTABLE RISKS 1 (1997) (“[O]rganizational and technological failures have become intimately linked so that to fully understand the cause of most major accidents, we must analyze both the administrative and technical aspects of the situation.”); PARKER, *supra* note 10, at 83 (“[T]o get at behaviour within the organization means changing the balance of external influences vying for management attention, so that social and legal responsibilities get a higher priority.”); DIANE VAUGHAN, THE CHALLENGER LAUNCH DECISION 67 (1996) (“Decision making in organizations is always affected by how information is sent and received, the characteristics of that information, and how it is interpreted by the individuals who send and receive it.”).

may have little to do with symbolism or belief. The character of a company or a trade union owes much to the structure of the industry, the skills of employees or members, the alliances that can be fashioned, and many other practical limits and opportunities. Attitudes and beliefs account for only part of an organization's distinctive character.

... "Character" refers to the commitments that help to determine the kinds of tasks an organization takes on, the opportunities it creates or closes off, the priorities it sets, and the abuses to which it is prone. . . . We cannot presume that every organization has a definite character. When one does, however, we can usually identify premises that fix, for substantial periods, the association's operative goals and characteristic methods.³⁶⁶

A key aspect of the corporate governance of financial firms—using this term broadly to include not only legal documents and roles but also attitudinal settings, discursive practices, unspoken norms, and patterns of social influence within firms—is the degree of attention that firm managers devote to deliberating on the possibilities of stress and failure.³⁶⁷ Managerial attention to these matters is determined in large part by patterns of social influence and the "intentional and unintentional usage of language to frame [issues]" within firms.³⁶⁸ These corporate governance systems are heterogeneous and subject to constant change.³⁶⁹ Some of these systems accommodate a meaningful deliberation on stress and failure better than others. With MBR, these systems become a proper subject for regulatory influence and control as the main structural pillar of an attempt to shape institutional character and industrial morality.

Once this point is admitted, then the logic of MBR emerges in clearer relief. One of the principal virtues of MBR is that it facilitates learning by regulators, who are able to intervene in that planning stage to identify the best practices of high-performing firms and present those practices prospectively as standards to lower-performing firms.³⁷⁰ It connects administrative law and regulation directly to the learning that occurs on

366. SELZNICK, *supra* note 362, at 321–22.

367. *Cf.* POWER, *supra* note 168, at 10 (arguing that corporate governance norms themselves are "a risk management strategy for a distinctive kind of risk—the failure of senior management to prevent risk incubation").

368. *See* Jane E. Dutton, *Strategic Agenda Building in Organizations*, in ORGANIZATIONAL DECISION MAKING 81, 89–90 (Zur Shapira ed., 1997).

369. *Cf. id.* at 86–99 (discussing the various intersecting forces that can affect agenda-setting by corporate decision makers).

370. *See supra* note 353 and accompanying text.

the ground at the industry level.³⁷¹ Again, the pragmatist influence is apparent.³⁷² Deliberation-oriented stress testing regulatory programs should be conceptualized as efforts to shape industrial morality, on a rolling basis, by encouraging firms to continuously evolve and improve the way they produce and process information concerning the risk of stress and failure. In this way, the extent of the corporate deliberation on safety and failure is no longer determined by any fixed statutory or regulatory standard and is instead set by a dynamic and dialogic interaction between industry and regulators over what are the *best* standards of failure awareness.³⁷³ Andrew Haldane of the Bank of England makes this point with regard to the regulation of stress testing:

Having banks conduct regular evaluations of their positions relative to a set of common scenarios (provided by the authorities) would be an improvement on current practices in several respects. First, it would allow some degree of benchmarking of results across institutions; second, it would allow a degree of benchmarking, and hence peer review, of models; and third, *it would hopefully help in ensuring stress-testing exercises form an input to management decisions and are not an annual regulatory ritual.*³⁷⁴

In this way, regulators can use information learned from industry to direct supervisory resources in a risk-responsive manner, focusing on bad actors and slow movers.³⁷⁵

371. See Dorf & Sabel, *supra* note 346, at 350–51.

372. See *supra* notes 353–54 and accompanying text.

373. Joseph Rees notes that, after Three Mile Island, nuclear power operators channeled their institutional resources away from compliance with Nuclear Regulatory Commission rules and toward achieving “standards of excellence.” See REES, *supra* note 14, at 1–2, 20–21.

374. Why Banks Failed the Stress Test, *supra* note 170, at 8 (emphasis added). There is some evidence that regulators were aware of this benefit of horizontal review, though their track record in harnessing the benefit has been poor. See *Lessons Learned in Risk Management Oversight at Federal Financial Regulators: Hearing Before the Subcomm. on Sec., Ins., & Inv. of the S. Comm. on Banking, Hous., & Urban Affairs*, 111th Cong. 14 (2009) (statement of Roger T. Cole, Director, Division of Banking Supervision and Regulation, Board of Governors of the Federal Reserve System) (“[T]here was a significant opportunity to put pressure on the big firms to improve their ability to pull positions together on a firm-wide basis and develop a really robust stress test. . . . We used that as a major tool in terms of pushing on those firms. It was feedback from that exercise and saying, look, you need to do more here, and that is one of the main tools that we have, is that type of horizontal review.”).

375. Cf. Robert Baldwin & Julia Black, *Really Responsive Risk-Based Regulation*, 32 LAW & POL’Y 181, 181 (2010) (“These are collections of strategies that, at the very least, involve the targeting of enforcement resources on the

There is a further, somewhat counterintuitive, advantage of these rolling best practices regimes: they also create economic incentives for those firms with institutional characters predisposed to promote regulatory objectives to innovate their practices. If, for instance, Bank of Prudence communicates to regulators a new stress test that, for the first time, takes seriously the possible linkages between leveraged loans and repo funding ability, regulators might insist that Bank of Hazard and other less mindful banks conduct the same test and consider the results' implications. Bank of Prudence indirectly initiates a conversation between the bank regulator and Bank of Hazard concerning risks that the latter might have been avoiding altogether, perhaps in the process taking market share away from Bank of Prudence and even contributing to systemic risk in the financial system.³⁷⁶

In performing this exercise, regulators are not working from a *tabula rasa*. Financial firms already have distinctive institutional characters within their risk management departments, upper echelons of management, and boards of directors. Mikes' field research into risk management practices at large financial firms, discussed above in Part II.B, is relevant here. She found that two prevailing "calculative cultures" predominate within financial firm risk management departments: quantitative enthusiasm and quantitative skepticism.³⁷⁷ Quantitative skeptics consider risk measurements as trend indicators to be taken into consideration alongside qualitative criteria such as "managerial discretion, experience and judgment."³⁷⁸ They use quantitative models as "learning tool[s]" in multifactor judgment processes rather than as an "answering machine."³⁷⁹ For them, "risk control is akin to a devil's advocate system, to be mobilized in order to challenge taken-for-granted assumptions and foster organizational learning."³⁸⁰

Mikes found further that the salience of the role of risk management in firm corporate governance depended in large part on the extent to which risk managers aligned their technical capabilities with the predominant calculative culture

basis of assessments of the risks that a regulated person or firm poses to the regulator's objectives.").

376. See Dorf & Sabel, *supra* note 346, at 350–51.

377. Mikes, *supra* note 56, at 22.

378. Mikes, *supra* note 57, at 15.

379. See Mikes, *supra* note 56, at 35–36.

380. *Id.* at 22.

within their respective firms.³⁸¹ In quantitatively skeptical firms, risk management personnel exert greater influence over agenda setting and key strategic decision making.³⁸² Such firms conceptualize “risk” in broad enough terms to include not simply measureable uncertainty but any threatening event irrespective of its ability to be measured.³⁸³ Because they had information—albeit at times non-quantifiable information—concerning such risks that senior management valued, risk managers were expected to discuss them.³⁸⁴ On the other hand, in quantitatively enthusiastic firms, risk managers were unable to impact big-picture strategy but were able to marshal quantitative risk estimates to influence already-existing debates about how to allocate capital in the firm.³⁸⁵ Quantitatively enthusiastic management demands quantifiable risk estimates as the key building blocks of an “economic capital” managerial system that adjusts the profitability of business lines according to the risks they posed.³⁸⁶ Senior management instructs risk managers to quantify risk so as to “induce correct economic behaviour” in light of the firm’s goals.³⁸⁷

From the perspective of stress test planning, not all calculative cultures are equal. There are several reasons why deliberation-oriented stress testing is more likely to flourish in quantitatively skeptical risk management departments. The first reason flows from the conceptual roots of stress testing in the study of failure.³⁸⁸ A firm’s orientation to failure is necessarily a strategic issue. On account of the correlation between quantitative skepticism and risk managers’ participation in strategic decision making, regulators should encourage quantitative skepticism. That is to say, risk managers, who are responsible for conceiving of and implementing stress testing programs, will have access to board-level attention. A second, but related, reason flows from the fact that deliberation-oriented stress testing requires an encounter with non-

381. *Id.* at 37. Mikes builds on Anish Bhimani’s earlier work finding that the perceived success of a management information system depends on whether the cultural premise of a new system is aligned with the predilections of the intended users of the new system approach. *Id.* at 21.

382. *Id.* at 28.

383. *Id.* at 35.

384. *Id.* at 29.

385. *Id.* at 34.

386. *Id.* at 24, 32.

387. *Id.* at 31.

388. *See supra* Part I.A.

quantifiable uncertainty that quantitative skeptics are better positioned to perform. A third reason is that by demanding precise calculation outputs, quantitative enthusiasts are inclined to orient towards assurance-oriented stress testing and furthermore make unavailable several of the types of stress testing exercises highlighted above, such as war games and reverse stress tests.

C. HIGH-RELIABILITY ORGANIZATIONS: REGULATORS ALREADY HAVE PROTOTYPES AND NEED NOT CREATE ANYTHING NEW

In using MBR to foster the quantitative skepticism conducive to deliberation-oriented stress testing, regulators have other institutional models from which to draw. In particular, management researchers have identified a category of so called “high reliability organizations” (HROs) that are characterized by their aspiration of complete failure avoidance and their commitment to resilient performance in conditions of uncertainty and stress.³⁸⁹ Since an MBR regulatory program to encourage deliberation-oriented stress testing touches directly on the concept of failure, the relevance of HROs in this context is readily apparent. Remember that the conceptual roots of stress testing and stress analysis are found in the search to better understand the network of causes out of which failure might result. In fact, we have already seen how one branch of engineering—systems reliability analysis—expressly directs itself to the concept of reliability.³⁹⁰ In the brief description of HROs that follows, bear in mind how HROs focus on *the causes of failure*.

Examples of HROs include nuclear power plants, aircraft carriers, wildfire-fighting crews, disease control authorities, and air traffic control systems.³⁹¹ The decisional infrastructures of these organizations prize reliable performance because failure—e.g., a nuclear power plant meltdown or a mid-air collision of commercial jetliners—is catastrophic from the perspective of managers.³⁹² Financial regulators therefore would do well to look to the institutional character-morality of HROs.³⁹³

389. See Rijpma, *supra* note 14, at 39.

390. See *supra* Part I.B.

391. WEICK & SUTCLIFFE, *supra* note 14, at ix.

392. See *id.* at ix–x.

393. It might be objected that the analogy to financial institutions is not complete because the tolerance of failure for financial institutions in general is low but not zero. Making banks completely failsafe would also ensure that the

Kathleen Sutcliffe and Karl Weick individuate the following five attributes of HROs: (1) they are preoccupied with the possibility of failure; (2) they are reluctant to accept simple explanations; (3) they maintain sensitivity to actual conditions of operations; (4) they are committed to resilient performance that maintains dynamic stability even in the presence of continuous stress; and (5) they are non-hierarchical and defer to expertise wherever it is located.³⁹⁴ The authors label the first three attributes “principles of anticipation” and the latter two attributes “principles of containment.”³⁹⁵

The three principles of anticipation are relevant to deliberation-oriented stress testing because stress analysis is at bottom a diagnostic tool—it identifies rather than fixes problems.³⁹⁶ Recall that risk management is about identifying threats to desired objectives and taking steps to control them; stress testing is only concerned with the former task.³⁹⁷ Sutcliffe and Weick discuss anticipation in terms with which scenario analysts at banks would be well familiar: “To anticipate is to foresee or imagine an eventual unchecked outcome, based on small disparities [between observations and expectations].”³⁹⁸ They further use the concept of *mindfulness* to illustrate how these principles contribute to the distinctive institutional character of HROs:

Mindfulness . . . involves the combination of ongoing scrutiny of existing expectations, continuous refinement and differentiation of expectations based on newer experiences, willingness and capability to invent new expectations that make sense of unprecedented events, a more nuanced appreciation of context and ways to deal with it, and identification of new dimensions of context that improve foresight and current functioning. . . . Mindfulness is focused on clear and detailed

banks would be of little use to savers and borrowers. *Cf.* COUNTERPARTY RM GROUP REPORT, *supra* note 180, at 11 (“[S]tress tests, when combined with carefully constructed scenario analyses, can be helpful, but even under the best of circumstances, stress tests can never anticipate how future events will unfold unless such tests are so extreme as to postulate outcomes that no level of capital or liquidity will provide protections against potential failure.”). A privatized system of finance necessarily entails that financial institutions engage in risk-taking and therefore will be subject to some non-trivial possibility of failure. Nevertheless, policymakers’ failure tolerance approaches zero for systemically significant financial institutions whose failure could cause the core functions of the financial system to collapse.

394. *See* WEICK & SUTCLIFFE, *supra* note 14, at 8–17 (discussing the attributes of HROs).

395. *See id.* at 63–64, 81–82.

396. *See id.* at 63–64 (summarizing the anticipation principles).

397. *See supra* notes 47–58 and accompanying text.

398. WEICK & SUTCLIFFE, *supra* note 14, at 45.

comprehension of emerging threats and on factors that interfere with such comprehension. Small failures have to be noticed (the principle of preoccupation with failure), and their distinctiveness must be retained rather than lost in a category (reluctance to simplify). People need to remain aware of ongoing operations if they want to notice nuances that could be symptoms of failure (sensitivity to operations).³⁹⁹

As discussed at length above, risk management departments use VaR techniques and stress tests to form expectations about risk, loss, and failure in an uncertain economic future.⁴⁰⁰ These expectations about risk largely determine how capital is allocated throughout the corporate group and also contribute to strategic agenda setting.⁴⁰¹ But Sutcliffe and Weick note that expectations are a “mixed blessing” because they create “blind spots,” by which they mean “belated recognition of unexpected, threatening events.”⁴⁰²

They note further that our ability to update our expectations in light of new information is frustrated by the “disqualification heuristic,” which describes our tendency to disqualify disconfirming information, highlight confirming information, and neglect information that contradicts convictions.⁴⁰³ The disqualification heuristic is related to the so-called “overconfidence bias,” “hindsight bias,” and “outcome bias.”⁴⁰⁴ The overconfidence bias describes our tendency to have inflated subjective perceptions of correctness.⁴⁰⁵ The hindsight bias refers to our tendency to overestimate the amount of information we deem relevant at the time we made a decision.⁴⁰⁶ The outcome bias describes our demonstrated tendency to evaluate events in ways that are anchored on observed outcomes.⁴⁰⁷ These deci-

399. *See id.* at 32–33.

400. *See supra* Part II.B.

401. *See supra* notes 381–87 and accompanying text.

402. WEICK & SUTCLIFFE, *supra* note 14, at 23.

403. *See* Lee Clarke, *The Disqualification Heuristic: When Do Organizations Misperceive Risk?*, 5 RES. SOC. PROBLEMS & PUB. POLY 289 (1993).

404. *Cf. id.* at 299–301 (discussing how the disqualification heuristic leads to biased risk assessment).

405. *See* Sarah Lichtenstein et al., *Calibration of Probabilities: The State of the Art to 1980*, in JUDGMENT UNDER UNCERTAINTY: HEURISTICS AND BIASES 306, 314 (Daniel Kahneman et al, eds., 1982) (“The most pervasive finding in recent research is that people are overconfident with general-knowledge items of moderate or extreme difficulty.”).

406. *See* Baruch Fischhoff, *Hindsight ≠ Foresight: The Effect of Outcome Knowledge on Judgment Under Uncertainty*, 1 J. EXPERIMENTAL PSYCHOL.: HUMAN PERCEPTION & PERFORMANCE 288 (1975).

407. *See* Jonathan Baron & John C. Hershey, *Outcome Bias in Decision Evaluation*, 54 J. PERSONALITY & SOC. PSYCHOL. 569, 569 (1988).

sional pathologies have obvious negative effects on institutional learning with the familiar result that key corporate decision makers overlook accumulating evidence of anomalies, overestimate the validity of their expectations, and even struggle to distinguish narrowly-avoided catastrophe from outright successes.⁴⁰⁸ Barry Turner and Nick Pidgeon's "disaster incubation theory" formalizes these decisional pathologies and shortcuts into an organizational theory: weak signals of disaster are ignored during a latent phase of "disaster incubation" before pushing systems into disaster modes.⁴⁰⁹

These decisional phenomena explain our demonstrated tendencies to normalize unanticipated deviations from our expectations—even when they may be weak signals of impending catastrophe. HROs counteract this tendency by seeking instead to continuously problematize and anomalize the unexpected.⁴¹⁰ For instance, the military and U.S. intelligence community use

408. See WEICK & SUTCLIFFE, *supra* note 14, at 26.

409. See BARRY A. TURNER & NICK F. PIDGEON, *MAN-MADE DISASTERS* (2d ed. 1997). Other researchers of disasters and failures object that any systematic effort to discover weak signals of impending failure is bound to fail for complex systems such as the financial system, and its constitutive institutions and their portfolios. See, e.g., CHARLES PERROW, *NORMAL ACCIDENTS: LIVING WITH HIGH-RISK TECHNOLOGIES* 4–5 (2d ed. 1999) (suggesting failure is inevitable in most complex systems); Weber, *supra* note 10, at 665–704 (explaining why the financial system is complex in light of, among other things, Perrow's normal accident theory). To these theorists, every complex system already carries the seeds of a failure—or, to use Perrow's phrase, a "normal accident"—that will come to pass in a manner that defies *ex ante* prediction. See PERROW, *supra* at 4–5 (discussing the unpredictability of failures in complex systems). The events that HRO theorists point out as harbingers of disaster are only signals in hindsight. See *id.* Notwithstanding the emphasis on HROs in this Section, this Article avoids weighing in on either side of that debate. I have argued elsewhere that the financial system is in fact vulnerable to normal accidents and that the most effective way to counteract those vulnerabilities is to embark on structural reform of the sector. See Weber, *supra* note 10, at 710 (suggesting structural reforms). Instead, the Article undertakes an analysis of the uses to which stress testing may be put and makes recommendations for their implementation as a regulatory matter. In other words, it argues that deliberation-oriented stress testing is more likely to be successful if regulators encourage firms to adopt it as a method inspired by HROs; it stops well short, however, of presenting deliberation-oriented stress testing as a panacea for financial instability.

410. See WEICK & SUTCLIFFE, *supra* note 14, at 41–42. A related discipline of near-miss incident management attempts to design a systematic program for recognizing near-misses as signs of operational failure and learning from them. See James R. Phimister et al., *Near-Miss Incident Management in the Chemical Process Industry*, 23 *RISK ANALYSIS* 445, 446–47 (2003) (proposing a framework consisting of identification, reporting, prioritization/distribution, causal analysis, solution identification, dissemination, and resolution).

a so-called “red team” in operations planning to do just that.⁴¹¹ The red team is responsible for playing devil’s advocate during planning sessions, directly challenging conventional wisdom and existing plans, using “what if” types of questions to find flaws in decisions, forecast, or intelligence information.⁴¹² In other words, they do the opposite of what Admiral Ugaki did in the war games leading up to the Battle of Midway, with catastrophic results for the Japanese fleet.⁴¹³

Other frequent reference points in the HRO literature are the linkages between the institutional character of the National Aeronautics and Space Administration (NASA) and the high-profile failures of the Challenger and Columbia space shuttles.⁴¹⁴ The post-accident commissions charged with investigating the disasters framed their reports as answering nearly identical queries: why did NASA go ahead with the launches despite knowledge of serious problems and despite concerns of engineers?⁴¹⁵ In other words, what inhibited NASA from learning about failure from the accumulation of weak signals of stress? This is an inquiry into how NASA’s institutional character lacked an appropriate commitment to mindfulness of the risks of failure.⁴¹⁶

Diane Vaughan’s study of the 1986 Challenger disaster highlights how NASA failed to learn from weak signals of the problem that ultimately resulted in the shuttle’s catastrophic failure.⁴¹⁷ The immediate cause of the explosion was a faulty design in the O-ring that sealed the joints of the solid rocket

411. See BREMMER & KEAT, *supra* note 123, at 25.

412. *See id.*

413. See *supra* note 133 and accompanying text for a discussion of how Ugaki introduced bias into the results of the game, thereby reducing its information value.

414. See generally VAUGHAN, *supra* note 365 (discussing the institutions at NASA and the space shuttle disasters).

415. See COLUMBIA ACCIDENT INVESTIGATION BD., REPORT VOLUME I, at 195 (2003), available at http://www.nasa.gov/columbia/home/CAIB_Vol1.html [hereinafter COLUMBIA REPORT] (also discussing the Challenger commission report).

416. *Id.* at 177 (“Many accident investigations make the same mistake in defining causes. They identify the widget that broke or malfunctioned, then locate the person most closely connected with the technical failure When causal chains are limited to technical flaws and individual failures, the ensuing responses aimed at preventing a similar event in the future are equally limited . . .”).

417. See VAUGHAN, *supra* note 365, at 124–43 (discussing the events leading up to the challenger explosion).

boosters.⁴¹⁸ The O-rings allowed leakage of hot propellant gas to breach the joint and eventually ignite the liquid hydrogen and oxygen in the external fuel tank.⁴¹⁹ Vaughan's study finds that evidence of the design errors had been discovered during test flights, discussed, and reported on by NASA contractor working groups.⁴²⁰ In their institutional response to this information, the working groups "normalized" the deviance of the experience from their expectations.⁴²¹ By *normalization of deviance* Vaughan refers to the process by which "behavior the work group first identified as technical deviation was subsequently reinterpreted as within the norm for acceptable joint performance, then finally officially labeled an acceptable risk."⁴²²

A similar normalization of deviance occurred in the lead-up to the disastrous 2003 Columbia shuttle mission. The technical cause of the loss of Columbia and its crew was a breach in the so-called "Thermal Protection System" on the leading edge of the left wing, caused by a piece of insulating foam that had separated from an external fuel tank shortly after launch and struck the shuttle's wing.⁴²³

During re-entry this breach in the Thermal Protection System "allowed superheated air to penetrate through the leading edge insulation and progressively melt the aluminum structure of the left wing, resulting in a weakening of the structure until increasing aerodynamic forces caused loss of control, failure of the wing, and breakup of the [shuttle]."⁴²⁴

Foam-shedding had occurred repeatedly in NASA space shuttle losses, and had been consistently labeled an "In-Flight Anomaly," which required a specific NASA organization to resolve the problem or convince NASA that it did not pose a threat to the crew.⁴²⁵ NASA managers would then remove the designation as corrective measures were taken, but the problem kept recurring.⁴²⁶ When a major foam-shedding event took

418. *Id.* at xi.

419. *Id.*

420. *See id.* at 119–25 (discussing NASA's work groups).

421. *Id.* at 65.

422. *Id.* Vaughan identifies a recurring five-stage process in the work group's processing of information about erosion of the O-ring: (1) signals of potential danger; (2) official act acknowledging escalated risk; (3) review of evidence; (4) official act indicating the normalization of deviance: accepting risk; and (5) shuttle launch. *Id.*

423. COLUMBIA REPORT, *supra* note 415, at 9.

424. *Id.*

425. *Id.* at 123.

426. *Id.* at 196.

place in 2002—a year before the fatal Columbia launch—NASA managers downgraded it to the status of an “action item.”⁴²⁷ The board charged with investigating the accident expressly tied NASA’s institutional failures to those failures in the lead-up to the earlier Challenger launch.⁴²⁸ The institutional similarities between the accidents are striking: “In all official engineering analyses and launch recommendations prior to the accidents, evidence that the design was not performing as expected was reinterpreted as acceptable and non-deviant, which diminished perceptions of risk throughout the agency.”⁴²⁹

Robin Dillon and Catherine Tinsley use the Columbia tragedy as an illustration of a more general phenomenon concerning why weak signals of failure, which they term “near misses,” are rarely treated as presages of future failure worthy of targeted further inquiry.⁴³⁰ In their view, NASA managers processed the information of foam-shedding in a manner that actually *decreased* their awareness of risk.⁴³¹ Consequently, that lower sense of perceived risk encouraged future riskier choices—i.e., to continue with the shuttle launch despite further revelations of risk.⁴³² “Hence, “NASA managers contemporaneously accounted for, but failed to learn from, near-misses.”⁴³³ So how is it possible that learning of weak signals of failure decreased rather than increased subjective risk perceptions? Dillon and Tinsley hypothesize that the explanation lies in the difference between propensities and dispositions.

The propensity of an event can be distinguished from its disposition by reference to the concept of the “close counterfac-

427. *Id.*

428. *See id.* at 195 (drawing parallels between the explosions). Overall, the 2003 commission report’s description of NASA’s institutional character reads like a tutorial on how *not* to be an HRO. It decries an unjustifiably optimistic safety culture (i.e., no preoccupation with failure), *id.* at 180, consistent preferences for oversimplified explanations (i.e., no reluctance to simplify), *id.* at 181, and a rigid and hierarchical organizational structure within the mission management team that impeded the communication of information (i.e., no deference to expertise). *Id.* at 192 (discussing experts’ reluctance to speak).

429. *Id.* at 196.

430. *See* Dillon & Tinsley, *supra* note 11, at 1425–26 (discussing the treatment of near misses).

431. *Id.* at 1426.

432. *Id.*; *see* Matthew N. McMullen & Keith D. Markman, *Downward Counterfactuals and Motivation: The Wake-Up Call and the Pangloss Effect*, 26 PERSONALITY & SOC. PSYCHOL. BULL. 575, 581–83 (2000) (describing this phenomenon as the “Pangloss effect”).

433. Dillon & Tinsley, *supra* note 11, at 1426.

tual.”⁴³⁴ A close counterfactual exists where a subject affirms that the counterfactual outcome “almost” occurred—instead of merely affirming the counterfactual outcome “could have” occurred.⁴³⁵ Disposition refers to the subjectively understood statistical probability of an event occurring based on prior rate information.⁴³⁶ Propensity, by contrast, refers to the subjective perception of probability supplemented by *event cues* that subjectively tie the counterfactual outcome causally to the actions taken.⁴³⁷ To illustrate, consider the following scenario developed by Daniel Kahneman and Carol Varey to illustrate this point: “At the end of a long game of chance, John could have won the whole pot if a die that he rolled showed a six. The die that he rolled was loaded to show six 80% of the time. John rolled it and it showed a two.”⁴³⁸ When research subjects were prompted with subjects the close counterfactual “The die almost rolled a six,” they rejected it decisively.⁴³⁹ However, they largely accepted the close counterfactual “John almost won the whole pot.”⁴⁴⁰ The results show how subjective understanding of probability depends in part on information that does not impact objective probability at all. The key factor is whether event cues situate the information according to a causal script.⁴⁴¹

Dillon and Tinsley apply the insight into the risk discussion by inquiring into how near-misses affect perceptions of risk.⁴⁴² Near-misses can prompt either serious deliberation and updated probability assessments or they can be celebrated as evidence corroborating the robustness and safety of the existing state of affairs.⁴⁴³ The researchers found strong support for the

434. See Daniel Kahneman & Carol A. Varey, *Propensities and Counterfactuals: The Loser That Almost Won*, 59 J. PERSONALITY & SOC. PSYCHOL. 1101, 1101 (1990).

435. *Id.*

436. *Id.* at 1102.

437. See *id.* at 1104–05 (distinguishing close counterfactuals from propensities).

438. *Id.* at 1104.

439. *Id.*

440. *Id.*

441. See *id.* Two key factors impacting the existence of event cues are proximity and decisiveness. See *id.* at 1106–07.

442. Dillon & Tinsley, *supra* note 11, at 1430.

443. See Dillon & Tinsley, *supra* note 11, at 1430; Adam J. Hirsch & Gregory Mitchell, *Law and Proximity*, 2008 U. ILL. L. REV. 557, 569–70 (“In situations where assimilation [of the close call to the narrowly-avoided bad outcome] is likely to induce fright, a close call . . . acts as a ‘wake up call.’ But, in those instances where it instead prompts contrast, and so induces relief, a close call can cause complacency and thus impair learning.”).

hypothesis that near-misses are interpreted in a favorable light, decrease perceived risk, and encourage risky behavior.⁴⁴⁴ They conjecture that the increased risk tolerance is due to the absence of event cues connecting the near-miss to proximal failure.⁴⁴⁵ People adjust downward their subjective perceptions of risk as they weave a causal script linking the near-miss to non-failure rather than to failure.⁴⁴⁶

The Dillon-Tinsley results are worrisome from the perspective of financial regulators. They also enhance the case for HRO principles. Given the highly competitive environment in which modern financial institutions operate, the tendency to normalize deviant events (i.e., weak signals of failure such as localized pockets of increasing loan delinquency or market rumors of liquidity concerns) results in a failure to relate the event to a causal script involving the possibility of failure. It is easy to imagine how this plays out in the construction of VaR models, sensitivity analysis, and stress scenarios: in the relative quiescence of present market conditions, market anomalies are normalized or, worse still, interpreted as signs of the robustness of the same quiescent conditions. This is a familiar story of how—despite the enormous resources devoted to developing the “sophisticated” risk management systems—clusters of weak signals of impending catastrophe become blind spots and go unheeded by risk managers, executives, and directors.

One of the central challenges in implementing an effective deliberation-oriented stress testing regime is therefore how to problematize deviations from expectations so as to enable them to, among other things, more readily construct causal scripts leading to failure. The commission that investigated the September 11, 2001 terrorist attacks impliedly invoked HRO principles in a discussion of how to prevent catastrophic attacks on major domestic infrastructure, insisting that governmental agencies “routiniz[e], even bureaucratiz[e], the exercise of imagination.”⁴⁴⁷

444. See Dillon & Tinsley, *supra* note 11, at 1431–36.

445. *Id.* at 1431.

446. See *id.* at 1433–34.

447. 9-11 COMM’N REPORT, *supra* note 16, at 344. One commissioner clarified that “you don’t really institutionalize imagination” so much as “you put in place systems that allow the imagination that’s naturally occurring to actually break through.” Jena McGregor, *The Gospels of Failure*, FAST COMPANY, <http://www.fastcompany.com/52512/gospels-failure> (last visited Apr. 21, 2014) (quoting Jamie Gorelick).

But what sort of imagination concerning stress should be encouraged? Research into the psychology of proximity has demonstrated that proximity is a decisive factor in determining how an event is processed: “as proximity to calamity or success increases, the likelihood of strong reactions to the ultimate outcome increases, counterfactual dwelling on ways the outcome could have been altered increases, and behavioral changes are more likely to follow, especially with respect to potentially repeatable outcomes.”⁴⁴⁸

Conceptually, deliberation-oriented stress testing should have greater success in achieving financial stability where risk managers experience failure as a psychologically proximate event. But that is just to re-state the problem: the key question from an MBR perspective is how to encourage such associations. Mindful HROs provide examples to which financial firms, prodded by regulators, should aspire.

One small note is in order concerning a large problem: that of incentives. So far, this discussion of HROs has elided considerations of economic motivations to cut expenses and grow revenue: devoting resources to the three principles of anticipation—attentiveness to weak signals of disaster as harbingers of failure, skepticism with respect to expectations and rules of thumb, and reluctance to simplify—is costly. The sort of mindful operations and planning culture that pervades HROs undercuts the usefulness of heuristics and routines that usually do the job right and save time and resources. Even more importantly, allowing for mindfulness and the principles of anticipation to control capital allocation in a diversified financial conglomerate is costly, especially considering the intense competition in the industry.

It should be noted that most of the organizations highlighted by the HRO literature are governmental units or utilities that receive some protection from market competition.⁴⁴⁹ And even with these entities, HRO management does not arise naturally. Rees has observed that only when the nuclear power

448. Hirsch & Mitchell, *supra* note 443, at 570.

449. See *supra* note 394. One notable private sector HRO noted by Sutcliffe and Weick is the courier FedEx’s Global Operations Control Unit, which manages a “sweep network” of FedEx airplanes. See WEICK & SUTCLIFFE, *supra* note 14, at 70–71. Every evening, twenty or so FedEx airplanes leave their departure points only 60% full so as to permit them to re-route in order to pick up unanticipated cargo and ensure that FedEx is able to meet its promised obligations without exceptions. Consequently, FedEx achieves added resiliency by maintaining slack in its operations. *Id.*

utilities realized in the aftermath of Three Mile Island incident that their future viability depended on industry-wide reliability—that they were “hostages of each other”—did an industrial morality develop to motivate a commitment to high reliability.⁴⁵⁰ It should not surprise that purely private-sector HROs are rare. One example of a private-sector, HRO-like operation is the Toyota production system (TPS).⁴⁵¹ This distinctive production system, long the subject of management studies, starts from a set of shared commitments to trust, zero tolerance for defects, continuous improvement (*kaizen*), and the indeterminate value of “quality” that is subject to continuous reevaluation in light of new circumstances.⁴⁵² In furtherance of these commitments, TPS espouses decentralized and collaborative decision making, deliberate destabilization of routines to facilitate learning, and a commitment to analyzing any failures to their root causes.⁴⁵³ When an unexpected problem occurs on the assembly line, workers have the power to pull the “*andon cord*” that halts production at the entire plant so that the anomaly can be examined and discussed.⁴⁵⁴

Though the parallels to HRO principles are evident, recent revelations of product defects and subsequent recalls call into question the ability of Toyota, or any private industrial concern, to remain committed to such principles in the face of pressures to cut costs and grow revenues and margins.⁴⁵⁵ These private sector realities underscore the need for regulators to take an active role not only in overseeing the scenarios, but also shaping the industrial morality. This will necessarily entail

450. See generally REES, *supra* note 14 (discussing how nuclear industry operations have changed since the Three Mile Island incident on account of more effective self-regulation).

451. See Simon, *supra* note 353, at 45–47.

452. See *id.*

453. See *id.* at 45–55.

454. *Id.* at 45.

455. Current Toyota CEO Akio Toyoda has stated that market pressures caused Toyota’s priorities to “bec[o]me confused,” leading Toyota to accelerate production too quickly for company engineers to ensure quality commitment as much of the work was outsourced to suppliers. See Alan Ohnsman et al., *Toyota Recall Crisis Said to Lie in Cost Cuts, Growth Ambitions*, BLOOMBERG, Feb. 26, 2010, <http://www.bloomberg.com/apps/news?pid=newsarchive&sid=aF0aX8t0Q6lk>.

heightened regulatory engagement during examinations with both risk managers and senior executives.⁴⁵⁶

All this is not to say that regulators should abandon the use of MBR regulatory tools to foster deliberation-oriented stress testing, but rather that in implementing such a program, they should be willing to see cost-cutting and imprudent use of leverage or expansion of revenue as part of the problem that needs solving. The results of these tests can be taken into consideration during the bank examination process.⁴⁵⁷ Although this Article stops well short of advocating for regulatory allocation of credit, a shift towards a more hands-on engagement of questions of risk assessment by regulators befits the regulation of an industry that, like the utilities that historically *have been* subject to regulatory price setting at the retail level, is responsible for key economic infrastructure without which the economy cannot properly function. It is time to take seriously the notion of the *utility function* of financial institutions.⁴⁵⁸

CONCLUSION

Although stress analysis is a parvenu in the bank regulatory regime, it has a long history in the engineering field from as early as the sixteenth century. These early stress testing methodologies evolved into professional norms on the part of engineers to remain focused on worst-case scenarios when de-

456. *Cf.* Why Banks Failed the Stress Test, *supra* note 170, at 9 (“In the arm-wrestle with management, it is about supplying power to the elbow of risk-managers.”).

457. U.S. bank regulators subject large banks to a continuous supervision regime. *See, e.g.*, OFFICE OF THE COMPTROLLER OF THE CURRENCY, COMPTROLLER’S HANDBOOK: LARGE BANK SUPERVISION 17–21 (2010) (explaining that examination of large banks involves a periodic core assessment that culminates in a report from the OCC to the bank’s board of directors as well as “various ongoing supervisory activities” and “targeted examinations”—i.e., integrated risk assessments by business or product line). Regulators usually rely on “the use of reason and moral suasion” as their “primary corrective tools.” FED. DEPOSIT. INS. CORP., RISK MANAGEMENT MANUAL OF EXAMINATION POLICIES § 15.1 (2010). The use of these soft persuasive tactics occurs under the shadow of bank regulators’ statutory powers (i) to order banks to remediate unsafe or unsound practices uncovered during examinations backed by specified and open-ended enforcement authority and (ii) to issue cease-and-desist orders with respect to unsafe or unsound practices. *See* 12 U.S.C. §§ 1818(b), 1831o (2012) (outlining the powers of bank regulators).

458. *See generally* John Kay, Narrow Banking: The Reform of Banking Regulation (Sept. 15, 2009) (unpublished manuscript), available at <http://www.johnkay.com/wp-content/uploads/2009/12/JK-Narrow-Banking.pdf> (drawing the parallels between the financial system and utilities).

signing and building structures, materials, and systems. Financial firms have adopted an extensive suite of stress testing techniques alongside their risk management systems. These techniques represent the most direct encounter with the concept of failure by and among firm decision makers. Most of the discussion regarding stress testing, however, focuses too much on the mix of stress testing tools a firm uses and not enough on the decisional infrastructure of, or orientation towards, stress testing. While financial regulators and policymakers are right to conceive of stress testing as a proper subject of regulation, their conceptual understanding has lagged. In particular, their efforts from the late 1980s until today have largely considered stress tests as verification tools rather than as part of a deliberative, mindful institutional orientation towards failure- and stress-related information within the corporate governance infrastructure of regulated firms. The success of stress testing regulation will depend in large part on whether regulators shift their focus towards this latter model.