What do stress tests test?

Citation for published version:

Digital Object Identifier (DOI):
10.1111/1468-4446.12739

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
British Journal of Sociology

Publisher Rights Statement:
This is the peer reviewed version of the following article: Coombs, N 2020, 'What do stress tests test? Experimentation, demonstration and the sociotechnical performance of regulatory science', British Journal of Sociology, pp. 1-17, which has been published in final form at: https://doi.org/10.1111/1468-4446.12739. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving.

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
What do stress tests test? Experimentation, demonstration and the sociotechnical performance of regulatory science

Nathan Coombs
School of Social and Political Science, University of Edinburgh, Edinburgh, UK

Correspondence
Nathan Coombs, School of Social and Political Science, 15a George Square, University of Edinburgh, UK
Email: nathan.coombs@ed.ac.uk

Abstract
After their successful introduction during the 2007-9 financial crisis central bank stress tests were adopted as a fixture of international banking supervision. However, in recent years a new normal has emerged where banks are expected to pass the tests, raising questions about the tests’ usefulness and legitimacy. Combining a dramaturgical interpretation of regulatory science with the idea of performativity in the sociology of finance, this article understands stress tests as a sociotechnical Goffmanian performance. With a focus on the Bank of England’s programme, the paper argues that the Bank’s decision to make their tests ‘predictable’ is an attempt to shore up central bank legitimacy by constraining regulatory discretion. This is accomplished through the use of calculative and procedural stage management techniques which allow the Bank to control the contingency of the testing process whilst demonstrating its objectivity. Nevertheless, the conclusion suggests that in the context of low levels of trust in central banks routine declarations of ‘all clear’ may undermine public confidence in the tests’ credibility and necessity. The study draws on 20 interviews with high-level regulators, financial practitioners and other stakeholders in the Bank of England’s stress tests.

Keywords: stress testing, central banks, finance, regulation, performances, legitimacy
1. INTRODUCTION: WHERE’S THE DRAMA GONE?

When in 2016 the Bank of England announced the results of their annual stress test, the UK’s Channel 4 News visualised the story by showing the Royal Bank of Scotland’s (RBS) logo straining as bricks piled on top. While the testing of building materials has little in common with the complicated calculative procedures involved in financial stress tests, the image helped to dramatize the results. It was not just that RBS’s capital would fail to stay above a relatively arbitrary regulatory minimum in a hypothetical scenario where unemployment shoots up to 12% and house prices fall by 35% (Bank of England 2016). The Scottish bank would implode, fracture, shatter irrecoverably. The image amplified the Bank’s message: RBS should raise more capital, urgently.

Such visualisations have become rarer in recent years as reports on stress tests have retreated from televised news programmes to the pages of the specialist financial press. For although in their early years a significant number of global banks were failed in stress testing programmes, both the Federal Reserve’s and Bank of England’s tests have since seen a steady decline in the number of failures, settling at zero at the time of writing (Figure 1). This does not mean that the announcement of a test’s results has become entirely drama free: a few banks continue to skim close to their regulatory capital minimum on some measures, or be criticised on other grounds. But with so few institutions being failed outright, the occasion has become predictable. In the language of the financial sector, the tests are now BAU: Business as Usual. All of which has led commentators to ask whether the tests have outlived their usefulness (Tett 2015).
Figure 1: Number of banks failing the stress tests (2012-19)

Source: Author's own based on Federal Reserve and Bank of England data. Failure in the Federal Reserve’s Comprehensive Capital Analysis and Review (CCAR) is defined by ‘objections to the capital plan’. For the Bank of England’s tests, failure is defined as a bank resubmitting their capital plan during the testing process.

What explains the transformation of stress tests from a show of regulatory force to a relatively uneventful ritual? Drawing on 20 interviews with high-level regulators, financial practitioners and other stakeholders in the Bank of England’s stress tests, this article argues that the Bank’s decision to frame their tests as ‘predictable’ was motivated by the need to manage financial sector concerns about regulatory discretion whilst reassuring the public that the financial system is secure. To understand that dual messaging the paper combines a performative interpretation of ‘regulatory science’ (Irwin et al. 1997; Jasanoff 1987, 1995, 1990; Rushefsky 1986; Salter, Leiss, and Levy 1988) with the notion of performativity in the sociology of finance (Callon 1998; Callon, Millo, and Muniesa 2007; Coombs 2016; MacKenzie and Millo 2003; Millo and MacKenzie 2009). Hilgartner’s (2000) Goffmanian understanding of regulatory performances is particularly relevant here, since it points to how the authority and credibility of policy-relevant science is not given but must be actively cultivated through persuasive frontstage demonstrations. Where the paper goes further is in showing how the experimental core of the testing process is also shaped by the communicative needs of the central bank. This involves the deployment of calculative and procedural stage management techniques to control the contingency of the testing
process and demonstrate its objectivity. In short, the paper shows that the predictability and low/no failure rate of recent tests is as much a sociotechnical accomplishment as the earlier use of stress testing to legitimise the recapitalisation of the banking sector.

At the same time, the paper’s critical coda contends that these technocratic procedures might have unintended and counterproductive consequences. Following commentators who argue that confidence in central banks has still not been restored after the financial crisis (Braun 2016; Dietsch, Claveau, and Fontan 2018; Riles 2018; Tucker 2018), it is possible to conjecture that repetitive judgements of ‘all clear’ may be received incredulously by their public audience or simply contribute to the tests losing their ‘hold on the collective imagination’ (Jasanoff 2005: 248). That has implications for their political economy: without an attentive audience the tests are more likely to be progressively watered down to accommodate financial interests, or dispensed with altogether. Ultimately, then, the wider message is that studies of high-profile public tests should be about ‘the sort of social and political relationships embedded within society as whole’ (Pinch, 1993, p. 38). With testing becoming an increasingly ubiquitous governmental technology, researchers should place the technical details of tests in the large-scale interpretative frame to understand how, when and under what conditions they are taken as legitimate forms of knowledge production on which to base policymaking.

The rest of the paper is organised as follows. The first section revisits key concepts in the sociology of testing and brings them into dialogue with work on regulatory science and central banking. The second section presents the study’s data sources. The third section interrogates how financial actors perceive the tests’ accuracy and the performative function of pass/fail judgements. The fourth section shows that the tests also serve as a vehicle for experimental ‘macroprudential’ policies and how they have elicited concern about regulatory discretion. The fifth section details three stage management techniques employed by the central bank to control the contingency of stress tests and their reception. The sixth section reflects on the potential implications of the low/no failure rate of recent tests for the legitimacy of the practice. The conclusion follows.
2. REGULATORY SCIENCE IN CENTRAL BANKING

Testing may not at first seem the most sociological of subjects. But as the nascent sociology of testing showed in the 1980s and early 1990s, once realist preoccupations are dispensed with it is possible to examine the different frames of meaning communities of practitioners bring to experimental results (Collins, 1988, 1987; Collins and Pinch, 1979; MacKenzie, 1990; Pinch, 1993). From this perspective, testing is not a merely technical exercise but a practice invested with interpretative and political saliency where the accuracy of results can be contested by social actors (MacKenzie, 1990). Addressing how this dynamic plays out on the public stage, Collins (1988) makes an important distinction between ‘experiments’ and ‘demonstrations’. Experiments are pre-closure attempts to infer the properties and behaviours of technologies among the ‘core-set’ of specialists. Demonstrations, on the other hand, are post-closure attempts by specialists to exhibit experimental results so that the public so that they can see ‘with their own eyes’. Collins’s examples are tests intended to demonstrate the safety of flasks of nuclear waste and a new type of jet fuel. Despite both tests having ‘apparently clear and convincing outcomes’ (Collins, 1988, p. 733), these were interpretations by journalists and onlookers which did not align with the judgements of the tests’ designers.

Coming at a similar set of issues but from an institutional level of analysis, political sociologists have grappled with the differences between ‘regulatory science’ and regular research science (Irwin et al. 1997; Jasanoff 1987, 1995, 1990; Rushefsky 1986; Salter, Leiss, and Levy 1988). Building on the observation that in the latter half of the twentieth century regulatory agencies were tasked with increasing responsibility for risk assessment that forced them to take an active role in knowledge production (Jasanoff 1990), the concept of regulatory science (and equivalents such as ‘mandated science’, ‘trans-science’ and ‘policy-relevant science’) draws attention to sociologically distinctive character of scientific research when entangled with policymaking. Scholarship has observed how the expectation that regulatory agencies respond rapidly to emerging risks pushes them to lower the evidentiary bar compared to academic research science while inviting adversarial challenge from businesses and civil society actors concerned that science is being manipulated to support an ideological agenda. Meanwhile, another strand of the literature has pointed to the problem of democratic accountability when governmental decision making become the responsibility of opaque technocratic committees, alienating publics from the agencies entrusted with representing their interests (Bijker, Bal, and Hendriks,
Hilgartner’s (2000) interpretation of regulatory science as a Goffmanian performance attempts to make sense of how regulatory actors navigate these dilemmas. On their back stage, regulators are well aware of the precariousness of the knowledge on which they have to make decisions. That is why the authority of science is mobilised so forcefully on their front stage: to suppress uncertainty and reassure the public that regulatory decision making is objective and credible.

Interestingly, central bank operations have not previously been analysed through the lens of regulatory science despite researchers drawing attention to the ‘scientization’ of central banking and its immersion in ‘technical rationality’ since the 1980s (Abolafia 2012; Marcussen 2009; Mudge and Vauchez 2016). Nonetheless, similar conclusions have been reached independently. Although central banks are preeminent producers of macroeconomic knowledge in certain academic fields (Dietsch, Claveau, and Fontan 2018; Mudge and Vauchez 2016), scholars have noted that this body of technical knowledge has been continually tested and contested. Goffman’s dramaturgical analytic has therefore been invoked to make sense of how central banks sustain public trust in their operations (Abolafia 2012; Braun 2016). For example, Abolafia notes that during the ‘stagflation’ crisis of the 1970s, which threw into question the use of the ‘Phillips Curve’ to steer interest rate policy and left rudderless the Federal Open Markets Committee’s private deliberations, in public the U.S. Federal Reserve continued to communicate complete confidence in its policy decisions (Abolafia 2012). Others have also noted that macroeconomic forecasts play a frontstage role in bolstering the credibility of the central banks’ policy commitments despite the poor track record of forecasts in predicting future developments (Beckert, 2016; Evans, 1997).

This paper argues that the Goffmanian notion of performance can also shed light on central banks stress tests but that their performative aspects should be located in the sociotechnical operations at the core of the experimental process. For as a large body of science studies inspired economic sociology has demonstrated, models and market devices play a vital evaluative and coordinating role in financial markets (e.g. Callon 1998; Callon, Millo, and Muniesa 2007; MacKenzie and Millo 2003; Millo and MacKenzie 2009). That is because they do not simply represent financial ‘objects’ but constitute them in the act of measurement. Less has been written on how regulatory authorities make use of models and devices (Coombs, 2016, 2017, 2020; Lenglet, 2011; Langley 2015; Williams 2009), but the point carries across: because the meaning of categories central to regulatory practice such as ‘risk’ and ‘value’ are sociotechnically constructed on the basis of radically reflexive and
uncertain market dynamics (Beckert and Bronk 2018), these categories are highly reactive to acts of measurement. Regulators are aware of these performative dynamics and stage manage their stress tests accordingly.

The article also offers new insights about how regulatory performances confer legitimacy on the decision making of governing authorities. Taking seriously the role played by publics in constructing the legitimacy of regulatory science, work on public reason has proposed the concept of ‘civic epistemologies’ (Miller, 2005, 2008; Jasanoff, 2005, 2011, 2012) to make sense of the culturally-specific ‘shared understandings about what credible claims should look like and how they ought to be articulated, represented and defended’ (Jasanoff, 2005, p. 249). The major insights of this work stem from a comparison between the United States, United Kingdom and Germany, which demonstrate a surprising diversity of ‘social and epistemic arrangements’ (Miller, 2008, p. 1898). For example, in the United States quantitative analysis is preferred and the legitimacy of regulatory science is established adversarially in the legal system, whereas in the United Kingdom trust is placed in public reports by respected civil servants. Yet the attachment of the concept of civic epistemology to national cultural comparison (Jasanoff, 2005, p. 270) has arguably played a role in closing off other research avenues, such as the historical and political context that shapes how publics receive and interpret regulatory performances. With a focus on the role played by stress tests in the post-financial crisis political conjuncture, this article provides an example of the idiosyncratic ways in which publics may be enlisted (or fail to be enlisted) in the construction of regulatory legitimacy.

3. DATA SOURCES

The paper draws on 20 semi-structured interviews (each lasting 45-150 minutes) with high-level regulators, financial practitioners and other stakeholders in the Bank of England’s (Bank’s) stress tests (22 interviewees in total). Launched in 2014, the Bank’s public testing programme is a relative newcomer compared to its international counterparts. The European Banking Authority’s (EBA) programme stretches back to the crisis-era Eurozone tests in 2009, and the Federal Reserve’s (Fed’s) most challenging programme (the Comprehensive Capital Analysis and Review – CCAR) has run annually since 2011. Perhaps because it was able to learn from these institutions’ experiences, the Bank’s programme is generally considered a compromise between their approaches. The Fed’s CCAR is a ‘top-
down’, labour-intensive process in which banks submit their balance sheet data to the Fed, who then runs the stress simulations. In contrast, while the EBA crafts the scenario its ‘bottom-up’ approach delegates responsibility for running simulations to the banks themselves and their national authorities.’

Because of the Bank’s evenly-weighted combination of top-down and bottom-up testing – in which the Bank sets the scenario, the commercial banks simulate it themselves, and the Bank checks their results – it was necessary to speak to both public and private actors involved in the process. These interviewees were contacted using a snowball method beginning with contacts established by attending three industry conferences on stress testing in London in 2016. The initial aim of the research was to discover how the stress scenarios are designed, interpreted and modelled. Over time, the investigation broadened to explore how the policy space has changed over the years, and comparisons between different central banks’ approaches to stress testing. The mix of public and private sector interviewees provided a balanced sample of interpretative and normative expert opinion on the motivations for, quality of, and potential problems with the Bank’s stress tests.

Given that this research concerns a confidential policy space, all interviews were conducted on the promise of anonymity. However, table 1 (below) breaks down the interviewees by professional category. In a number of cases, where individuals had moved between positions and across the public-private divide over preceding years, they are classified according to their most relevant professional position. Most of the key individuals at the Bank involved in the designing and implementing its financial stability architecture and stress testing programme were consulted. So too were senior personnel involved in managing risk, stress testing and capital allocation at the UK’s commercial banks. It should also be noted that some regulatory interviewees requested to read the manuscript prior to publication. They used this opportunity to correct technical errata but did not comment on its arguments or substantive claims.

Further sources which inform this discussion include national and transnational policy documents on central bank stress testing, commentaries in the financial press, and reports by industry groups and think tanks. Finally, a review of the archive (1997-2018) of the banking sector’s premier trade journal, Risk, was conducted to understand the historical background of the technique.
| Current and former regulators at Bank of England and European regulatory institutions | 8 |
| Current or former risk, stress testing and treasury managers at UK banks | 8 |
| Financial ‘quants’ and software engineers | 3 |
| Other stakeholders (consultant, lobbyist, academic) | 3 |

Table 1: Study’s 22 interviewees (2016-19) listed by professional category

4. PERFORMATIVE MEASUREMENT

As the sociology of testing has observed, while scientists and the public at large see testing as experimental if it improves the accuracy of knowledge, it is possible to understand testing as socially productive without abiding by realist commitments – testing can be seen as producing useful knowledge even if one suspends judgement on whether the ‘truth’ is being approached asymptotically. In his analysis of the US Treasury’s 2009 Supervisory Capital Assessment Programme, Langley (2013) makes just that move: he interprets the test’s success in delivering a ‘positive affective charge’ to markets during the turmoil of the financial crisis as a function of its perceived ‘precision’ rather than its supposed accuracy. This section goes further by introducing a Goffmanian dimension to the analysis (Goffman, 1959). It argues that while the central bank’s pass/fail judgements play a performative role in potentiating supervisory interventions they also suppress uncertainty about the accuracy of the measurement on the experimental back stage of the process.

Although the origins of stress testing tend to be associated with state-directed interventions during the financial crisis (Geithner 2014), it has for a long time been a specialist, subaltern risk management technique in the financial sector. For example, in the absence of recent experiences of nationwide mortgage defaults in the United States, in the 1980s mortgage-backed securitizations were stress tested against the macroeconomic conditions of the Great Depression to estimate their risk (MacKenzie, 2011). Stress testing became more prominent in banking in the 1990s and 2000s when transnational capital regulations insisted that banks’ use the method alongside price-based statistical techniques
such as ‘Value-at-Risk’ (Izquierdo, 2001; Quagliariello, 2009). However, it was not until the 2007-9 financial crisis, and the breakdown of banks’ risk management systems, that regulators took control of the process in their supervisory interventions. Rather than letting each bank create its own scenarios, regulators crafted a common scenario and evaluated the results comparatively across the banking system, allowing them to direct state recapitalization efforts and release useful information to market participants.

Given the perceived success of these exercises, the procedure was rapidly instituted as a regularised fixture of post-crisis international banking supervision. There are procedural differences between different countries’ testing regimes, but a test invariably begins when the central bank releases its annual scenario – a spreadsheet of severely adverse variables projected 3-5 years into the future (these typically include macroeconomic indicators such as GDP and unemployment, as well as financial variables such as inflation and interest rates). In the Bank’s programme, responsibility then passes to the commercial banks to expand these variables into a more detailed macroeconomic projection, estimate the probability that their loans will default and the losses incurred, and calculate the impact on their capital. The results are then returned to the central bank along with a detailed narrative account justifying their assumptions, choice of models and governance process (the ‘qualitative’ component of the test). In the event that a bank dips below the regulatory capital minimum in the scenario simulation, the central bank will likely require it to raise more capital or impose restrictions on its dividend payments and share buybacks (what the Bank euphemistically calls ‘resubmitting the capital plan’).

The stated aim of the tests is to assess banks’ ‘capital adequacy’ by exposing them to ‘severe but plausible’ hypothetical crisis scenarios and bringing their capitalization levels up to a sufficiently resilient standard (Bank of England, 2015). While the results of a decade of central bank led testing have yet to be put to the test in an actual crisis, the history of the technique does not impart great confidence in the accuracy of the tests’ results. Critics have noted that there is no shortage of examples of where stress tests painted a flattering picture of the strength of financial institutions just prior to their collapse (Dowd, 2015a, 2015b). Indeed, like all forms of financial risk management there are serious questions to be asked about their accuracy (Millo and MacKenzie, 2009). The insight holds particularly in respect to commercial banks’ own pre-crisis stress tests, which by all accounts were remarkably untesting and failed to leave a meaningful impression on their risk management or conduct (Interview, 13 September 2016; Risk, 2006). Yet if the notion of ‘experimentation’ is unmoored from concerns with accuracy and attached to a
performative understanding of acts of measurement as productive of new realities (Callon, 1998), then the relatively large numbers of banks that have failed the regulatory tests over the years’ suggests that the procedure can be considered experimental – they led to supervisory actions that would not otherwise have been possible. There is also evidence that the tests have furnished knowledge enabling major interventions by policymakers. For example, according to the testimony of Andrew Bailey, then Deputy Governor of the Bank of England, the plug was pulled on a planned deal between the Co-operative Bank and Lloyds in 2013 only after the results of the stress test came in (Interview, 22 November 2016).

Another important consideration is how regulators and financial practitioners understand the knowledge produced by the tests. While there is still a tendency in public discourse and certain sociological quarters to see financiers as ‘model dopes’ (MacKenzie and Spears, 2014, p. 419) convinced that they can transform uncertainty into tidy quanta of risk, heightened concern with ‘model risk’ in the aftermath of the financial crisis has meant that many regulators and practitioners have internalised a more pragmatic stance. As a stress testing manager at a major UK bank put it: ‘Whether something is 10 or 20, it doesn’t really matter… the one thing about stress testing and forecasting, the answer is wrong. It won’t ever be that’ (Interview, 22 November 2016). Stress tests may not provide the ‘truth’ of how a bank would really fare in a doomsday scenario, but they do ‘tell you the direction where things move, how things are interacting, where risk may come about. And that’s the important bit’ (Interview, 22 November 2016).

From the regulatory standpoint, the tests are valuable because they ‘create a lot of information for the banks themselves and for their supervisors’ (Interview, 14 December 2016). Or as another regulatory interviewee put it: ‘None of the supervisors care that much about the headline number that comes out of the exercise. It’s getting the firms to do the analysis which is interesting’ (Interview, 22 November 2016). That might be an exaggeration given that it is the ‘headline number’ which determines whether banks pass or fail the tests. But it does point to how, for the actors involved in the testing process, belief in the accuracy of the results is not a prerequisite for seeing the exercises as productive experiments. In addition to serving a performative role in legitimating supervisory interventions, the pass/fail judgements provide the frontstage ‘expressive equipment’ (Goffman 1959: 22) of the experimental process: imparting a sense of accuracy and certainty that specialists, aware of the complex backstage apparatus of supervisory requirements and calculations, tend to be sceptical about.
4. MACROPRUDENTIAL EXPERIMENTATION

We have seen that the measurement of capital adequacy is interpreted more critically on backstage than it is presented on the public frontstage. However, there is an even deeper and hard-to-penetrate backstage of the stress tests reserved for a ‘core-set’ (Collins, 1988) of technical experts. As the Bank states when describing the purposes of their programme, it ‘is not solely about calculating estimates of bank capital in the adverse scenario. Rather, it represents a set of tools that allows policymakers to explore and better understand the vulnerabilities of the financial system’ (Bank of England, 2015, p. 9). In that exploratory capacity, stress tests conform closely to the understanding of regulatory science as pushing beyond the cognitive scope of regular research science (Irwin et al., 1997; Jasanoff, 2011). What distinguishes ‘macroprudential’ applications of stress tests are just how far to the outer edge of regulatory knowledge they skirt. With these functions lending a potentially high level of discretionary power to the central bank, this section argues that the Bank’s decision to frame their stress tests as ‘predictable’ can be understood as a performance of accountability intended to help establish the tests’ legitimacy in the eyes of financial practitioners.

For almost a decade before the crisis, thinkers at the Bank for International Settlements (BIS) in Basel Switzerland (the central bank for central banks) had been sounding the alarm about the limits of ‘microprudential’ supervisory approaches (Borio, 2003; Clement, 2010). Instead of focusing regulatory efforts on ensuring the sound management of individual banking institutions, they instead proposed a ‘macroprudential’ approach for safeguarding financial stability. In the aftermath of the crisis, with policy entrepreneurs from the BIS exerting greater influence, this led to an ‘ideational shift’ (Baker 2013) that invited central banks to adopt a ‘god’s eye view’ of finance, encouraging them to assume a more ‘commanding stance’ (Dorn, 2016, p. 1). How this ambition should be put into practice was, however, left open-ended and indeterminate (Stellinga, 2019). While initially articulated in bold anti-cyclical terms, where regulators would seek to smooth the credit cycle and deflate speculative bubbles, in many jurisdictions this gave way to the less ambitious resilience agenda focused on just increasing the quality and quantity of bank’s loss absorbing capital (Thiemann, 2019). Uniquely, the Bank of England was able to combine both agendas. After receiving enhanced statutory powers from Parliament, the Bank employed ‘strategic ambiguity’ (Best, 2012; Van Gunten, 2017) and subordinated ‘anti-cyclical policies to the goal of resilience’ (Thiemann, 2019, p. 570). At the same time, the
architect of the Bank’s new approach to financial stability governance, Paul Tucker, insisted that resilience was non-quantifiable and would be a moving target based solely on the judgement of the newly-formed Financial Policy Committee (FPC) (Kynaston, 2017, p. 775). On that basis the Bank was able to present its anti-cyclical policies as motivated by and contributing to the overarching resilience agenda.

The way that stress tests sit at the ‘junction of macro and micro prudential supervision’ (Interview, 22 November 2016) means that they are well-placed to perform this goal. One of the ironies of the macroprudential ‘paradigm shift’ is that while often framed by authorities and understood by scholars as an all-encompassing reorientation towards the systemic level, most of the tools at regulators’ disposal remain rather traditional. The increasing space the shift has afforded for blue skies approaches to financial system modelling (such as agent-based and network approaches) has not been matched by broader channels for regulatory intervention (which continue to operationalised mainly through microprudential supervisory channels). From the central bank’s perspective, the appeal of stress testing is that it has ‘two functions’ (Anderson, 2016, p. 11): assisting in the project of macroprudential knowledge creation while being connected to supervisory interventions for governing individual banks. Hence, from 2016 onwards, the Bank infused macroprudential anti-cyclical rationales into its essentially microprudential stress tests with the introduction of the ‘Annual Cyclical Scenario’ (ACS). The scenarios’ severity would now reflect the judgement of the FPC about the current position in the financial cycle (Bank of England, 2015, p. 5). Further, the tests could ‘inform’ the setting of countercyclical capital buffers, in which minimum capital requirements would be notched up and down (from 0-2.5%). So, while from 2016 the tests as a whole were given an anti-cyclical justification, the countercyclical buffer would allow the FPC, who meet on a quarterly basis, to make immediate interventions in response to emerging threats.

Given the novelty of the countercyclical buffer, private sector interviewees were unsure about its merits. At one institution, though, an individual responsible for a large UK bank feared that rather than reflecting objectively the position in the financial cycle the buffer could just reflect whatever the Financial Policy Committee ‘has got a bee in its bonnet about’ (Interview, 14 December 2016). This interviewee was concerned that the buffer is used ‘in the way that its presented in its policy statement to Parliament… and not to just have unlimited power to impose huge capital surcharges on banks at the drop of a hat’ (Interview, 14 December 2016). These remarks reflect a more widespread unease, expressed in feedback to the Bank’s discussion paper on its stress testing programme.
(Bank of England, 2014), that the Committee’s power to judge the position in the financial cycle might stray into unacceptably discretionary territory, providing regulators with an ‘unfettered ability to control individual and system-wide bank capital ratios’ (Bank of England, 2014, p. 7). That reflects unease at the political level, expressed as far back as 2011, with the discretion granted to the FPC with its macroprudential mandate (Baker 2013).

The Bank’s decision to frame its stress tests from 2016 onwards as ‘predictable’ (Bank of England 2015: 25) should be understood in this context. The notion of predictability is well-established in modern monetary policy and is integral to idea of governing interest rates by shaping market expectations (Blinder, 2004; Holmes, 2014; Wansleben, 2018). In much the same way that market participants are supposed to be able to anticipate movements in interest rates by attending closely to policymakers’ signals, the idea is that market participants can predict developments in the stress scenarios’ severity and decisions about the use of the countercyclical buffer by consulting the FPC’s financial stability reports (Brazier, 2016, p. 76). In the words of a regulatory interviewee: ‘We know that it takes time for firms to build capital. So to the extent that we can make the scenarios more systematic and predictable, it can help firms with that’ (Interview, 23 August 2016). Yet given the common association of stress testing with ‘thinking the unthinkable’ and criticism that the tests have become predictable over the years (Glasserman and Tangirala, 2015), it is surprising that the Bank adopted this language. To be clear, the predictability that the Bank publically endorses does not necessarily imply a zero failure rate regime. However, as the next section will show, this framing is not unrelated to the sociotechnical stage management of the tests that produces these results.

What then explains the Bank’s adoption of the potentially confusing notion of ‘predictable’ stress tests? This paper cannot present a definitive account, but the sociology of public knowledge provides clues when it emphasises how sensitive regulators and regulated are to perceived transgressions of objectivity and accountability (Jasanoff, 2011). To be accused of abusing their discretionary license and pulled into public controversy is a point of acute sensitivity for central banks who cherish their operational independence and justify it based on their value-neutral expertise (Conti-Brown, 2016). Those fears are heightened in the case of the Bank’s macroprudential regulation given how far it pushes beyond the economic knowledge sanctioned by academic research communities (Thiemann, Aldegwy, and Ibrocevic, 2017; Thiemann, 2019). The financial cycle that the Bank’s scenarios and countercyclical buffer are anchored to remains an uncertain theoretical postulate, contingent to a large extent upon the measurement device applied
to financial and macroeconomic data (Kranke and Yarrow, 2018; Stellinga, 2019). A regulator at the Bank conceded: ‘there’s never going to be a single financial cycle, it depends on what approach you take to try and estimate it’ (Interview, 15 March 2019). Given the degree of uncertainty and the level of regulatory discretion permitted by the macroprudential policy mandate, it is therefore credible to see the Bank’s embrace of predictability as intended to reassure financial market participants that their discretionary license will not be exercised liberally. Accountability is performed to their private sector audience in a rhetorical framing of stress tests at odds with common understandings of the practice.

5. THREE ACTS OF STAGE MANAGEMENT

The previous sections detailed two experimental applications of stress tests. First, as a performative measurement technique potentiating supervisory interventions. Second, as a vehicle for pursuing macroprudential policies at the outer limits of the central bank’s regulatory science. This section addresses the stage management techniques employed by the central bank so that these experiments enhance rather than undermine ‘public confidence in the banking system’ (Bank of England, 2015, p. 9). While the goal of boosting confidence seems in tension with the idea of an authentic experiment – even implying a propagandistic imperative which undermines the tests’ objectivity (Dowd 2015b) – the ambiguity of the measurement standards creates room for manoeuvre. For as the Bank insists, there is no objective standard of resilience than can be fixed and quantified. Neither is there a ‘correct’ level at which stress scenario severity would be sufficiently testing of banks’ resilience: ‘severe but plausible’ is vague enough to encompass stress events from the mundane to the extreme. The central bank thus has considerable latitude to employ calculative and procedural techniques to manage the testing process and to attempt to reconcile its experimental and demonstrative goals.

5.1 Scenario modelling and impact anticipation

A striking aspect of the Bank’s tests of capital adequacy is the extent to which the results fall within a narrow range. Some banks come close to failing, or fail marginally, but there are very few instances of banks failing spectacularly or cruising through the test unscathed. This is surprising: one might expect that tests exploring ‘severe but plausible’ doomsday
scenarios would produce volatile results. Even with the embrace of ‘predictability’, the consistent performance of commercial banks in the exercises requires explanation. To understand this regularity requires stepping back stage. Years before the launch of the Bank’s public stress testing programme the Bank developed their own ‘top down’ stress testing model called the Risk Assessment Model of Systemic Institutions (RAMSI). In contrast to ‘bottom up’ tests, where banks use their own models and utilise their detailed balance sheet data when simulating the scenario, RAMSI makes use of banks’ publically available balance sheet data (Burrows, Learmonth, and McKeown 2012). While reliance on this public data reduces the granularity and precision of the ‘top down’ stress test, it does not blunt RAMSI’s usefulness for shaping the testing procedure. First, at the preparatory stage, when experimenting with potential scenarios. Second, to anticipate the chosen scenario’s impact on individual banks’ balance sheets. Third, to check that banks’ results do not divert too far from what the Bank expected them to be.

RAMSI is therefore a vital technology for the Bank to fine tune the choice and ‘calibration’ of its scenarios in meeting its policy objectives. One of these objectives involves setting the scenarios’ severity at a level that reflects the FPC’s judgement about the current position in the financial cycle. Another objective revealed by this research involves pushing banks to the edge of failure without actually failing them. A regulatory interviewee close the scenario design process put it this way:

So I think there is a genuine emphasis on, sort of, almost trying to get the right scenario, which ultimately is a scenario, in the [Financial Policy] Committee’s judgement, they’d be uncomfortable if banks failed with it... They are sort of saying, if there is a scenario much worse than this, you know, you might expect some banks to fail. We are not a zero-failure regime, but they are trying to get to that level.

(Interview, 23 August 2016)

Since the differences between the projections produced by RAMSI and banks’ own simulations are of a degree rather than of a kind, the Bank can attach a probability to a bank failing the test prior to a bank conducting its own test. An interviewee at a large UK institution revealed that one of their primary interests prior to running a stress tests is in the Bank informing them of the likelihood of a ‘capital shortfall actually materialising’
(Interview, 14 December 2016). By this the interviewee does not mean the probability of an extreme event actually occurring, but the probability of their bank failing the stress test and being asked to raise more capital. These comments suggest that the trend of banks passing the stress test is a policy choice deliberately enacted through the central bank’s modelling technologies. A stress testing manager at a large UK bank confirmed this intuition: ‘frankly no one is going to fail the stress test from a quantitative perspective... [with] the type of scenario the PRA [Bank of England’s Prudential Regulation Authority] is using’ (Interview, 22 November 2016).

5.2 Foreclosing weakness and demonstrating strength

For obvious reasons, banks are not keen on failing the stress test: it can result in them having to raise more capital and carries ‘reputational risks’ that the market can punish them for (in share price devaluations). Central banks also have an interest in a low number of banks failing their stress tests; they are attuned to the reactive nature of markets, where an authoritative declaration of a bank’s weakness can turn into a self-fulfilling prophesy if depositors withdraw their money or funding markets turn away their business (Interview, 30 October 2016). Central banks are sensitive to the ways in which bank failures might reflect poorly upon their supervision: if a bank fails a test questions might be asked about why they passed it in previous years and what has gone wrong in the interim. Finally, regulators are receptive to the opinions of the banks themselves, who do not like the pass-fail judgements and lobbied against them.

One option for the central bank to prevent failures is just to stop issuing these judgements (as did the European Banking Authority in 2017). The Bank’s approach, in keeping with the preference of the British Bankers’ Association (BBA and BSA, 2014, p. 14), has instead been to imply failure but to avoid using the word explicitly. Whereas in the Federal Reserve’s reports ‘objections to the capital plan’ is a proxy for ‘failure’, there is no clear equivalent in the Bank’s reports, which only identify which banks were required to resubmit their capital plan or raise more capital. Further ambiguity derives from the lag between the deadline for banks returning their stress test results to the central bank in the spring and the central bank’s publication of the results in the winter. In the interim, there is ample time, if a bank performed poorly, for it to raise more capital and for the central bank to spin the results into a more positive narrative. One regulatory interviewee insisted that ‘you want to be in control of the messaging to the media’ (Interview, 30 October 2016).
Another described this as ‘supervisory actions behind the scenes’, citing the case of Standard Chartered’s poor performance in the 2015 test:

So Standard Chartered last year [2015] were a bit too close to the edge... So in the gap between July and December they went and raised capital. When the report came out, you could get the story: Standard Chartered didn’t have quite enough capital at the start of the process, but they have now, which is a nice outcome for the central bank... you don’t really want people going: “oh my god, there’s a problem with a big bank”. You know that’s kind of the exact opposite of what you’re meant to be doing as a regulator’ (Interview, 22 November 2016)

An interviewee at another institution agreed that these supervisory actions should be the story rather than talk of passing or failing the test. Going further, they interpreted these actions as meaning that such outcomes have already been superseded within the Bank’s framework.

We don’t like the fail word; it’s not about failure or not failure; it’s about what information the stress test tells us and therefore what action the individual bank is taking with the supervisor... so in that sense you could say that we don’t have a hurdle rate [pass/fail threshold] ... we have, you know, being asked to resubmit the capital plan. (Interview, 14 December 2016)

This resistance to the pass/fail judgements was echoed by a banking executive who complained that ‘as soon as you start to label pass or fail, that’s a bit of a red flag... how can you pass or fail a theoretical question about a theoretical scenario about something that may or may not happen’ (Interview, 15 March 2019). The trend of increasingly few banks failing the Bank of England’s stress tests is therefore not just because the scenarios are designed to deliver that result. It is also a product of a retroactive procedural technique in which the central bank can emphasise the results from the perspective of the supervisory actions that have since improved a bank’s capitalization.
5.3 ‘A little extra loop’: Applying a gloss of objectivity

Within the Bank’s stress testing programme the countercyclical buffer operates as the rapid response vehicle for the tests’ general anti-cyclical framework. One of its distinguishing features is that unlike the stress scenario’s severity (which is determined by the judgement of the FPC), the decision on whether to apply the buffer is supposed to be ‘informed’ by the results of the stress test. Given that the scenarios are being designed in such a way that few banks fail the test, how can the tests furnish knowledge which helps policymakers decide at what level to set the countercyclical buffer? An individual responsible for a large UK bank doubted they could:

It still seems a little bit circular to me in that it seems to start from a perspective of saying: well we think that given where we are in the financial cycle we’re about a third of the way up and so banks should be holding about 1% of the counter-cyclical buffer, let’s design the scenario that stresses the banks to the point where the expected result is that they will have to hold a 1% counter-cyclical buffer (Interview, 14 December 2016).

Reflecting on the circularity and seeming redundancy of the procedure, the interviewee further argued that ‘inserting this stress test piece, it’s like a little extra loop... the same judgement would be there but instead of going straight from a to b, you a – stress test loop – b’ (Interview, 14 December 2016). This raises the question of whether ‘the stress test is actually telling you anything new, or is it telling you that yes it was possible for you to make the data fit what you originally believed to be the truth’ (Interview, 14 December 2016). Another banking executive saw the reasons given for the application of the countercyclical buffer as cynically ambiguous. The result is that the regulator can say about the stress test’s role in setting the countercyclical buffer: “‘oh it’s just informing us, we’re just using this as a judgement, oh it’s okay’. They just hide behind it [the ambiguity] because that gives them carte blanche to do anything they want to do’ (Interview, 15 March 2019).

If these insinuations are true, they suggest that in setting the countercyclical buffer the stress test simply provides a gloss of objectivity to the subjective judgement of the Financial Policy Committee. Or in the terms offered by the sociology of organizations, the testing device, on this account, serves a ceremonial function is sustaining the myth that the
central bank’s decision making is grounded in fully rationalised technoscientific knowledge (Meyer and Rowan, 1977). As scholars have noted, such performative scientism is common in central banking (Abolafia, 2012; Marcussen, 2009; Braun, 2015; Mudge and Vauchez, 2016). While in their operations ‘custom, compromise, and pragmatism mix with expert judgement’, the resulting confusion and uncertainty is ‘strategically obscured from public view’ (Abolafia, 2012, p. 94). At the same time, this insight does not mean that the central bank is harbouring ulterior rationales for applying the countercyclical buffer. If the financial cycle is as an artefact of measurement devices and expert judgement, then the stress test can be understood as a performative front stage device which legitimates the central bank’s interventions.

6. A PREDICTABLE ACT OF PUBLIC THEATRE?

The previous sections showed that the Bank has ample tools at its disposal to stage manage the testing process, and this is responsible, at least in part, for enacting the policy decision to have few banks fail the test. How are these increasingly predictable demonstrations of financial stability being received by the public? It is not possible to ascertain empirically what the public thinks of stress tests given that there have been no surveys of public opinion. Hence, the arguments which follow are somewhat speculative in nature. But by building on public commentaries that criticise the tests as a manipulative performance, it is possible to conjecture reasonably that recent developments may threaten the tests’ legitimacy in the public’s eyes.

Andrew Haldane, Chief Economist at the Bank of England, argues: ‘Trust is the lifeblood of all things monetary and financial, including central banks. And incredulity is Kryptonite’ (Haldane, 2017, p. 2). For that reason, the years following the 2007-9 financial crisis have been testing for central banks. Central banks were widely held to have been either complicit in the financial excesses of the preceding decade or guilty of fatal complacency (Johnson and Kwak, 2010). That had led commentators to argue that public trust in central banks has collapsed, with these institutions finding themselves on the receiving end of populist provocations, challenges to their independence, and scepticism about their policy commitments (Goodhart and Lastra, 2018; Riles, 2018; Tucker, 2018). In this context, does a zero failure rate in recent stress tests, and the implication that the
financial system is resilient to any plausible scenario, risk undermining the credibility of the procedure?

Critical commentaries suggest ambient scepticism about the tests among experts in the public sphere. For example, a report by the US Office for Financial Research on the Federal Reserve's testing programme found a remarkable degree of consistency in bank losses year on year, suggesting ‘increasing predictability over time’ (Glasserman and Tangirala, 2015, p. 16). The authors worry that this could lower the informational value of the results and ‘may lead to pressures to weaken the process, given the costs involved in its implementation’ (Glasserman and Tangirala, 2015, p. 19). Picking up on the report, the Financial Times columnist Gillian Tett goes further using the occasion to ask whether the tests have ‘outlived their usefulness’ (Tett, 2015, n.p.). Tett proposes a ‘cynical view’ and asks her readership to ‘accept that the real value of the tests is as a piece of public theatre’ (Tett, 2015, n.p.). In a series of influential reports for the Adam Smith Institute, Kevin Dowd casts doubt on the Bank of England’s stress testing programme for similar reasons: ‘even if the Bank had severe doubts about the strength of the banking system, it cannot admit to them... The stress tests cannot then be credible, because only a reassuring answer can ever be allowed’ (Dowd, 2015b, p. 27). Finally, the public interest advocacy group, Finance Watch, has accused the Eurozone stress tests of ‘trying too hard to reassure the public’ and argued that this undermined their intention to ‘dispel lingering doubts and restore trust’ (Stiefmüller, 2018, n.p.).

Even if these expert concerns do not diffuse to the wider public sphere there is still the risk that the tests’ lack of drama and surprise will cause the tests to recede from public view. As Jasanoff argues, in order to maintain their hold on the collective imagination ‘the utility of the state’s knowledge producing endeavours must repeatedly be brought home’ (Jasanoff, 2005, p. 248). Given the high-stakes political economy of banking regulation, that observation has particular saliency. It may be the case that we have reached the point at which central banks wish to disassociate stress testing with crisis governance and establish a new form of ‘peace time’ stress testing (Schuermann 2016). It may also be the case that banks are now as capitalised as regulators wish to see them and so continued failures would lack an obvious supervisory response. Nevertheless, it is still possible to imagine reconfigurations of the practice that would not see it converted into an uneventful, technocratic ritual. For example, banks could have their resilience and response to the scenarios awarded a positional grade on a rank from strongest to weakest. Or banks could be failed on a broader set of criteria. In any event, central bank stress tests become just a
routine conformation of the banking sector’s health the tests seem unlikely to maintain public interest. A consultant for the banking industry argues: ‘There becomes a danger that these things become routine and they become boring and dull and they’re just done for the sake of doing it’ (Interview 15 March 2019). From this perspective, predictability imperils the tests’ front stage communicative function of the tests:

Behind the scenes, all of those rules, it’s all quite technical. In actual fact, the man in the street doesn’t really get that, doesn’t really understand that… so what the regulators have been doing is that stress testing for me is a more political animal which they can more clearly articulate and by then putting a pass or failure to it, they can really turn the screw (Interview, 15 March 2019).

Given that capital regulation has a large effect on banks’ share prices (Goodhart 2015; Admati and Hellwig 2013), it is no small matter for the regulator to ‘turn the screw’. What happens when the screw is released? One potential future is for stress tests to be progressively watered down to accommodate the preferences of financial sector actors. Indeed, such changes are already afoot. In 2019 the Federal Reserve announced that fewer banks will be subject to the qualitative aspects of its testing programme and that it will be dropping pass/fail judgements on this component of the tests. The Bank’s stress tests are framed more ambiguously, making it harder to identify specific changes which reveal a loosening of supervisory stringency. But the decision in 2018 to repeat the same scenario as in 2017 confirms the increasing emphasis placed on predictability rather than testing banks with new and challenging crisis scenarios (Bank of England 2018).

Albeit inconclusive, these reflections point to the need for the sociology of testing to incorporate large scale interpretative frames into their analyses of regulatory performances. This paper only speculates about the how the historical and political context in which stress tests are performed shape their reception and legitimacy. But in a time in which public trials and experiments are becoming an increasingly ubiquitous governmental technology (Adkins and Ylöstalo, 2018), there is an opportunity for sociologists to add a further layer to their research by studying under what conditions test become a legitimate and credible basis for policymaking. To some extent this recommendation is anticipated by Pinch (1993). He makes a programmatic call for sociology of testing to not just document
the interpretative flexibility of different social actors, or the epistemic cleavages between experts and publics, but to situate tests within their broader social and political contexts. If this study has shown one thing it is that there are deep entanglements between the technical details of tests and the audiences to which they are performed which mean that any analytical prism focused on a single ‘scale’ will be partial at best.

7. CONCLUSION

Combining Goffman’s notion of social performances with an understanding of the performativity of economic measurement devices, this article finds that central bank stress tests are sociotechnically stage managed to control the contingency of the results and demonstrate their objectivity. The Bank of England’s decision to make its annual tests ‘predictable’ relies on an existing apparatus of calculative and procedural techniques that were already utilised when the central bank was more willing to fail commercial banks and make high profile supervisory interventions. At the same time, these observations are not meant to imply that stress tests are nothing more than a performance in the colloquial sense of the term. Even though the frontstage of regulatory stress testing has become uneventful, it is possible that the data collected in the exercises is proving useful for backstage macroprudential experimentation.

And yet, tests of this scale do not exist in a social vacuum. This paper concluded by asking whether stress tests, which are still immensely time consuming and expensive to administer, can survive their transformation into an uneventful ritual. Already the regulatory pendulum is beginning to swing back in a deregulatory direction. Does the lack of drama and surprise in stress tests risk losing the public’s attention and hence the popular support necessary to uphold supervisory stringency in the face of a pushback by powerful and well-funded financial actors? These dynamics might seem closer to the concerns of political economists than to the preoccupations of the science studies scholars associated with the sociology of testing. But if the sociology of testing is to grapple with the proliferation of trials and experiments as governmental technologies then such questions are unavoidable. If they can provide compelling answers then a revived sociology of testing might move from being a sub-field of the sociology of scientific knowledge to the forefront of economic and political sociology.
Acknowledgements

This article has benefited from feedback by Benjamin Braun, Dylan Cassar, David Stark, Bart Stellinga, Matthias Thiemann, Arjen van der Heide, and Timo Walter as well as the participants of the workshop ‘Put to the Test’ workshop held in London, December 2018 and the panel ‘Financial Governance II’ at Intersections of Finance and Society, London, December 2019. I also thank the three anonymous reviewers, whose incisive comments helped to improve the manuscript. Some of the interviews this paper draws from were funded by a grant from the Leverhulme Trust [ECF-2014-304].

References


The European Central Bank has taken a greater role in the process since introduction of the Single Supervisory Mechanism in 2014.