Economising Failure and Assembling a Failure Regime

Kurunmäki, Liisa; Mennicken, Andrea; Miller, Peter

Economiising Failure and Assembling a Failure Regime

Year: 2023

Version: Published version

Copyright: © 2023 the Authors

Rights: CC BY-NC-ND 4.0

Rights url: https://creativecommons.org/licenses/by-nc-nd/4.0/

Please cite the original version:

Economising Failure and Assembling a Failure Regime

Liisa Kurunmäki, Andrea Mennicken, and Peter Miller

CC BY-NC-ND 4.0
DOI: 10.4324/9780429355950-15

The funders: University of Jyväskylä, Finland.
Introduction

Sociologists have largely neglected the topic of failure, and particularly the economising of failure. Likewise, they have paid insufficient attention to the increasingly elaborate failure regimes that assess failing and pronounce on failure, and the ways in which such regimes support the social relations and institutions that structure economic behaviour within modern markets (Halliday and Carruthers 1996). This is notwithstanding the many adjacent literatures that have developed in recent years. These include the substantial sociological literature on economising (e.g. Çalışkan and Callon 2009; 2010; MacKenzie, Muniesa, and Siu 2007; Miller and Power 2013; Muniesa et al. 2017), the “New Public Management” and “Audit Society” literatures (Hood 1991; Humphrey, Miller, and Scapens 1993; Pollitt 1993; Pollitt and Bouckaert 2011; Power 1994; 1997), as well as those writings that have highlighted the ways in which economic ideas and instruments facilitate the “governing” of economic life (Cutler, Hindess, Hirst, and Hussain 1978; Hopwood and Miller 1994; Miller and O’Leary 1987; Miller and Rose 1990; Thompson 1986).

Our aim, in line with the volume overall, is to redress a neglected aspect of failure in the sociological literature – namely the economising of failure – and to do so in three stages. First, we consider the neglect of the topic of failure in sociology, and articulate the distinction between failure and failing, as well as the dynamic process of the economising of failure. Second, we examine briefly the economising of the economy through the economising of failure, the assembling of a failure regime for the corporate world across more or less the whole of the nineteenth century and the first half of the twentieth century. Third, we examine the economising of the public sphere, the way in which the notion of failure as an economic event in the corporate world came to travel into the public sphere, and particularly the domain of hospital-based healthcare in England, across the first two decades of the twenty-first century. For during the past two decades or so, a particular idea of failure together with its associated calculative infrastructure have been proposed not just for hospitals but also for schools, social work, prisons, universities, and much more. Attending to the category of failure focuses our attention on the calculative infrastructure and ideas on which they depend. It also directs our attention to the co-construction of the entities to be regulated and the bodies that are to regulate them to avert
underperformance or failure. We consider in the next section the neglect of the topic of failure in sociology and outline the features of our approach.

**Failure Regimes, Failing, and the Economising of Failure**

This neglect of the economising of failure is all the more puzzling as an economised category of failure now saturates public life. The deficit is more than empirical, for failure defined as exit from the market game takes us to the heart of economising and the phenomenon that has been dubbed neoliberalism (Brown 2015; Dardot and Laval 2013; Davies 2014). Failure defined as financial failure, leading to system exit, has come to be viewed by its proponents as not only inevitable but desirable, insofar as it is held to promote both competition and accountability. As the reach of market-based principles is expanded, so too is the scope of government through a vast apparatus of regulatory intervention, often and ironically in the name of increasing competition. This tension is critical, for without a relatively orderly regime for exit, market principles are ultimately unable to operate. Yet allowing or facilitating the possibility of exit for service providers in the public sphere typically goes hand in hand with the imperative to maintain services. Attending to the category of failure thus focuses our attention on the often overlooked “how” of economising (Çalışkan and Callon 2009; Kurunmäki, Mennicken, and Miller 2016; Miller and Power 2013).

Malpas and Wickham (1995) have commented on this curious sociological lacuna, which is all the more puzzling given the extent to which failure is intrinsic to the larger “projects” within which they argue social life takes place (Malpas and Wickham 1995, 39). Consistent with the arguments of Hunt and Wickham (1994), they have called for a Foucauldian sociology of failure that views attempts at control, including control through market mechanisms and other processes of economising which are constitutive of social life, as always failing, always falling short of their targets. Rose and Miller argued in very similar terms, depicting government, including the governing of economic life, as a “congenitally failing” operation (Miller and Rose 1990; Rose and Miller 1992, 190). Government, they suggested, is a problematising activity, for the ideals of government are intrinsically linked to the problems around which it circulates, the failings it seeks to rectify. It is around these difficulties and failures that “programmes” of government are elaborated, which are fuelled by the constant registration of failure, the discrepancy between aspiration and outcome (Rose and Miller 1992, 190).

There are of course exceptions to this neglect, but these tend to be relatively discrete contributions rather than a cumulative body of sociological literature focusing on the economising of failure. Halliday and Carruthers (1996; 2007; 2009) have provided what is perhaps the most sustained contribution in this respect, offering a socio-legal approach to bankruptcy law-making. Accounting researchers have also addressed the topic, and in so doing have highlighted the links between economising and the regimes that seek to identify and regulate both failing and failure, in the corporate sphere and in the public sphere more broadly (Kurunmäki, Mennicken, and Miller 2018; 2019; Kurunmäki and Miller 2013; Miller and Power 1995). Earlier contributions from organisational scholars identified this important topic, although there was little follow-up subsequently from that field (see for instance Meyer and Zucker 1989; Whetten 1980).

Historians of bankruptcy, particularly with regard to the United States, have offered detailed descriptions of the emergence of bankruptcy law. The early work of Warren (1935) charted the development of bankruptcy legislation in the United States, from the late eighteenth century until the Bankruptcy Act of 1898, based primarily on congressional debates. Several decades later, Coleman (1974) provided a careful study of state law-making on issues relating to insolvency. Jackson (1986) and Sullivan, Warren, and Westbrook (1989) offered differing yet similarly
normative assessments of bankruptcy law. More recently, Balleisen (2001) considered the economic and social meanings of bankruptcy in Antebellum America, with particular attention to the 1841 Federal Bankruptcy Act and its place within the shifting character of American capitalism, while Skeel (2001) charted the birth of US insolvency law across most of the nineteenth century, culminating in the enactment of the Bankruptcy Act of 1898 and subsequent transformations over the following century or so. 2

It fell to a pair of sociologists to undertake a more analytic study of how and why American bankruptcy law took its distinctive shape (Carruthers and Halliday 1998). Focusing on the process that led to the 1978 Bankruptcy Code in the United States, together with corporate law reforms enacted in England in 1986, they combined neo-institutional insights with an emphasis on the recursive relation between law and organisations, and the role that professions play in the process. But it is Sandage’s Born Losers: A History of Failure in America (2005) that comes closest to the analysis proposed here of the economising of failure. We consider this in more detail in the following section, with particular attention to what it tells us about the economising of the economy over the course of the nineteenth century. For now, we turn our attention to the broad parameters of our approach.

Our approach to the economising of failure is consistent with the calls noted above for a Foucauldian approach to the analysis of failure, although rather than focusing on the generic characteristics of projects or programmes, the inevitable gaps that arise between aspirations and outcomes, we focus on the emergence and assembling of specific failure regimes, and their intrinsic links with the phenomenon that has been dubbed neoliberalism (Brown 2015; Dardot and Laval 2013; Davies 2014). We draw attention to the fundamental tension between expanding the scope of market-based principles while also expanding the reach of government through regulatory intervention. And we highlight the challenges faced in devising failure regimes, whether across the nineteenth century for the corporate sphere or the last two decades for the public sphere.

A number of features define our approach. First, we argue that failure has none of the objectivity or inevitability often attributed to it. This is the case equally in the corporate sphere and the public sphere. One can of course chart the volume of corporate failures, and such failures can clearly have devastating consequences for individuals, families, and even entire towns or regions. But, important as such aspects are, failure is much more than a statistical event or personal experience. And it is also much more than a matter of profitability or asset strength, or indeed the inverse. The moment of failure is much more complex than “realist” appeals to underlying economic reality suggest. The calculative technologies of accounting that are so central to such assessments are themselves enmeshed in an assemblage of expert claims, modes of judgement, financial norms, political negotiations, and much more (Miller and Power 1995). The moment of actual failure only exists within this assemblage of actors, instruments, ideas, and interventions, even if the form that the assemblage takes varies from country to country, particularly with regard to the differing territorial demarcations between lawyers and accountants.

Second, and notwithstanding the importance of avoiding “realist” narratives of failure, it is equally important to avoid juxtaposing realists and constructivists as if this all too convenient dichotomy advanced our understanding of failure (Kurunmäki and Miller 2013; Latour 1993). Consistent with the previous point, we need a language that helps us understand the variations in the stability of all those practices and entities that emerge out of the assemblage of actors, instruments, ideas, and interventions that seek to operationalise failure. To put this differently, we view failure as an archetypical variable ontology object. If the actual moment of failure emerges within and through an assemblage of calculative practices, expert claims and pronouncements, legal procedures, financial norms and risk assessments, political judgements, and so forth, we
Economising Failure and Assembling a Failure Regime

need to be much more attentive to the gradations of possibility, the gradations in the stability of entities, and the varying degrees of possibility in the ability of an agent or network of agents to bring about the moment of failure. To view failure as a variable ontology object means attending to the multitude of interactions among all the components of the failure assemblage, and their gradual stabilisation. A whole host of actors and instruments have to be brought into play and have to achieve a significant degree of stability before the moment of failure can be pronounced. Before that can happen, there is an open-ended yet not limitless set of negotiations and interpretations (on the “dialectic of failure,” see Power 1997). The more these are stabilised, the more real the possibility of failure becomes, until the moment that failure is pronounced. That moment is the outcome of this multiplicity of components interacting rather than a brute reality that somehow imposes itself upon us in an unmediated form.

Third, we argue that the analysis of failure regimes requires consideration of both the ideas and instruments that make them possible. We use the term “calculative infrastructures” to designate how such assemblages intertwine the operational and ideational dimensions of governing failure, how they transform the very concept of failure, and how it is to be acted upon (Kurunmäki, Mennicken, and Miller 2019). Such infrastructures not only make ideas about markets and economic rationality operable, but they also animate and shape economic thinking itself, including the ideas of actorhood that are involved. Calculative infrastructures are inherently relational phenomena, for it is not a matter of starting from the study of a given object such as “failure,” but of analysing the sets of practices that form and fashion the idea of failure itself, and in such a way that it can become the correlate of a historically specific set of practices for acting upon it. And here our emphasis on the variable ontological status of such phenomena deserves re-emphasis, insofar as it is the stabilising of a chain of calculative instruments and ideas that makes it possible for failure to be acted upon by those entities given responsibility for such matters.

Fourth, and finally, we argue that it is important to distinguish between failure and failing (Kurunmäki and Miller 2013). Again, we emphasise that “failing” is not a given or objective state of the world, but is a status that is internal to regulatory and policy discourses, and is itself subject to multiple processes of interpretation, judgement, and intervention. Failing can be a protracted process with no inevitable outcome, and certainly no inevitability that actual failure will ensue. Interventions directed at those entities deemed to be failing may delay further decline and avert actual failure. Whether in the corporate sphere or the public sphere, the notion of failing goes together with a whole set of instruments for assessing performance, comparing it with that of others, deciding what problems exist, and evaluating the interventions that may be possible. Failing is about prediction rather than pronouncement; it is about the future rather than the past. In contrast to actual failure, failing has a residually optimistic dimension, insofar as it allows for the possibility of cure.

Economising the Economy

Economising the public sphere has a long prehistory. Economising the economy through the category and calculation of failure was an even more protracted process. For even in the corporate sphere, the economising of failure required a fundamental shift in how failure was understood, how it could be “forgiven,” and how the act of forgiving could be made operable through a relatively stabilised failure regime, as Sandage and others have so perceptively shown (Balleisen 2001; Mann 2002; Sandage 2005). In the early and mid-nineteenth century, failure was deeply personal, encapsulated in the term “loser” and other associated terms (Sandage 2005, 11 ff.). This notion of failure brought together the economics of capitalism and the economics of personhood. The various attempts on both sides of the Atlantic during the nineteenth century to figure
out whether and how to forgive failure were not only economic and legal matters but also profoundly cultural (Mann 2002). The redefinition of insolvency as arising from risk rather than sin entailed an acknowledgement that the vicissitudes of capitalism could lead to personal failure even despite hard work (Sandage 2005, 15).

This redefinition of failure as economic rather than moral was a key part in the forming of a liberal economy. At its heart was a new economics of selfhood, which over time would come to tally with the economics of capitalism (Sandage 2005, 12). In the United States, the passing on the same day in 1867 of the Bankruptcy Act and the Reconstruction Act felicitously paired the birth of failure with the birth of freedom. 3

Or, as Sandage (2005, 223) puts it, “liberty and slavery” gave way to a new measure of human worth: “success and failure.” With this step, the economic domain was fully constituted qua economic domain. This required more than a transformation of the idea of failure. It required the forming of an entire calculative infrastructure, a complex chain of calculative practices and their associated rationales for predicting and pronouncing failure. This economising of the idea of failure, and the forming of a reciprocally related chain of calculations including financial statements, ratio analysis, risk indexes, and credit ratings made failure calculable in the decades following the 1867 Bankruptcy Act. Corporate failure defined as an economic event emerged within and through an assemblage of calculative practices, financial norms, legal procedures, expert claims, and modes of judgement. With this transformation of failure went the correlate that, if failure could be forgiven, then re-entry into the market game could be permitted, a positive step according to the proponents of bankruptcy legislation.

This assembling of a failure regime for the corporate sphere across the nineteenth century entailed an important distinction between failure and failing, the latter an often protracted process with no necessary end point. Failing today is often paired with interventions directed at the entities in question, so as to delay further decline and avert failure. Despite its negative connotations, there is a residual optimism inherent in the notion of failing. A large array of devices makes it operable, including comparisons with the performance of others, assessments of the severity of the problems, and a range of options for attempting to mitigate or avert the problems, including refinancing, forced disposals, closure of segments of the entity, government support, and so on.

The above, together with our own research (Kurunmäki and Miller 2013), leads us to identify three distinct steps in the assembling of a failure regime for the corporate sphere in the United Kingdom and the United States across the nineteenth century and the first half of the twentieth century. First, the forgiving of failure through repeated attempts to enact enduring bankruptcy legislation across much of the nineteenth century. Second, the emergence and growth of credit rating agencies from the 1840s onwards, whose work largely consisted for the rest of the century in accumulating narrative accounts of the character of individuals. Third, the development in the last decade of the nineteenth century and the first half of the twentieth century of a calculative infrastructure for seeking to forecast failure, using financial ratios and risk indexes.

The first step in the economising of failure and the assembling of a failure regime for the corporate sphere was the “forgiving” of failure through bankruptcy legislation. In both the United States and the United Kingdom, a remarkably protracted and faltering series of legislative moves resulted in more or less stabilised bankruptcy legislation at around the same time. In the United States, the short-lived Bankruptcy Act of 1800 began the process, even if it was repealed after only three years. There was regular debate about replacing it over the following decades, but it was not until the Bankruptcy Act of 1841 that further federal legislation was passed. Importantly, this Act introduced for the first time the principle of voluntary bankruptcy, and covered all individuals, not just merchants and traders. This Act lasted little more than a year, and had been under attack even before it came into force. Following this, there was a gap of a quarter of a century
Economising Failure and Assembling a Failure Regime

before Congress enacted another bankruptcy law. The 1867 Bankruptcy Act, which lasted just over a decade, was a further important step in the economising of failure and the assembling of a failure regime, for it meant that economic failure no longer deprived individuals of their capacity to transact. Failure was no longer indelibly inscribed in the character of an individual, but was a more circumscribed economic phenomenon. The relatively enduring Bankruptcy Act of 1898 affirmed the economising of failure, ensuring that success and failure could be viewed as two sides of the liberal ideal (Sandage 2005, 223).

In England, the trajectory was broadly similar, even though the exact chronology and resulting systems were somewhat different. The Bankruptcy Act 1831 established a Court of Bankruptcy in London and created a new figure, the official assignee, which radically altered the institutional and legal framework for dealing with insolvency. The Winding-Up Act of 1848 regulated the control of liquidations and made the appointment of a public accountant more or less essential. The Bankruptcy Act of 1861 introduced the important principle of voluntary insolvency, and the Companies Act 1862 – often referred to as the “accountant’s friend” – established the position of “official liquidator,” while the Debtors Act 1869 reduced the ability of the courts to detain those in debt, even though some provisions remained. Henceforth, imprisonment in England was to be reserved for the punishment of crime, and not misfortune in trade (Di Martino 2005, 27). The ensuing 1883 Bankruptcy Act remained in place for just over a century and provided the framework for the administration of bankruptcy until the passing of the Insolvency Acts of 1985 and 1986.

The second important step in the assembling of a failure regime for the corporate sphere was the rapid growth of credit rating agencies from the 1840s onwards (Carruthers 2013). The forgiving of failure through insolvency legislation was thereby paired with the forecasting of failure through the establishment of a vast information infrastructure for both narrating and rating failing. The formation of Lewis Tappan’s Mercantile Agency began this process in 1841, when it opened its doors in lower Manhattan. In the words of Henry Thoreau, it was to become a “kind of intelligence office for the whole country” (cited in Sandage 2005, 99). This information infrastructure was to match the recently established physical infrastructures of telegraphy, railroads, and steamboats through a network of local informants who, instead of payment, would receive a portion of any debt collected from local defaulters. Within five years of its opening, the company had enlisted 679 informants, which reached 2,000 by 1851. This kept 30 clerks busy, receiving on an average day 600 new or updated reports and answering 400 enquiries. Much of the time of these credit reporters was spent chatting with traders, bankers, sheriffs, and tavern keepers, as financial statements were rarely obtained and trade information was negligible. Assessments were largely in terms of the character of the individuals concerned, with “bad egg” being typical of the terms used in the credit reports that made failure indelible. As Sandage (2005, 130) remarks, Americans had not yet learned by the 1860s to think of each other simply as numbers. Initial attempts to standardise such reports began in 1869, with the Bradstreet Company sending instructions to its reporters to itemise length of time in business, amount of own capital in the business, estimated net worth after liabilities, and so forth (Kurunmäki and Miller 2013, 1105–6). By 1880, R.G. Dun and the Bradstreet Agency had become a clear duopoly in the field of national credit reporting, and by the end of the nineteenth century, most manufacturers and wholesalers used the information they provided. That said, and notwithstanding the full-time reporters that covered some of the more densely populated areas of the United States, little had changed between 1865 and 1890 in other areas, which continued to rely on the reports of local “informants.”

The third step in the assembling of a failure regime for the corporate sphere was the separating of rating and narrating, the calculating of the probability of failure instead of assessing
it on the basis of rumour and personal intuition (Cohen and Carruthers 2014). The growing institutional distance and impersonality between capital providers and businesses, especially in the United States, paved the way for the development of new analytical knowledge capable of assessing corporate health and the solvency of borrowers. The year 1890 can be taken as a turning point in this respect, for in that year credit man Peter Earling from Chicago published a manual titled *Whom to Trust: A Practical Treatise on Mercantile Credits* (1890). Earling (1890, 13) called for a “better understanding of the ‘Science of Credits’” and sought to identify “the causes that lead to success or failure.” Just a few years later, in 1896, a national group called the National Association of Credit Men (NACM) was formed and was incorporated in New York State the following year. It attracted 600 members to begin with, which trebled during the first year and had grown to 33,000 by 1920. The NACM sought to transform the practice of credit assessment by defining it as an economic domain susceptible to financial calculations. The NACM called for a standardising and economising of the information on which credit reports were based. Subsequently, they joined forces with the American Bankers Association to support calls for audited financial statements to be made more widely available, and for credit reports to be prepared on the basis of standardised statements. By 1899, they were able to report that 133 firms were using these forms.

Before long, a chain of calculations was beginning to form that would provide the calculative infrastructure for assessing and seeking to predict failure. It was through this infrastructure that the distinction between failing and failure was solidified. The increased availability and uniformity of financial information that began to occur from the late 1890s allowed a set of “second order” calculations (Power 2004) – accounting ratios – to be produced. The “current ratio” (current assets divided by current liabilities) began to gain acceptance at this time, as it appeared to allow creditors to predict the likelihood that a firm would be unable to meet payments and would therefore fail. Banks increasingly began to rely on this ratio as a basis for approving loans, and the “50% rule” was widely recommended, meaning that a borrower’s current liabilities should not exceed 50% of current assets. The NACM’s Bulletin set out similar guidelines in 1902, and ratio analysis gained momentum rapidly.

Gradually, across the first two decades of the twentieth century, the notion that failure was calculable and probabilistic began to take hold. In 1905, Cannon, a pioneer of financial statement analysis, spoke of the “rules of the credit science,” and in 1919 Alexander Wall published a highly influential study in the Federal Reserve Bulletin titled *Study of Credit Barometrics* (see Cannon 1905; Wall 1919). Wall computed seven different ratios for 981 firms, stratified by industry and location, and detected significant variation, which was a departure from the then-customary usage of a single ratio. There was an avalanche of publications on ratio analysis during the 1920s, which continued unabated during the 1930s, notwithstanding criticisms that they did not portray fundamental relationships within the business, and that the pursuit of the perfect set of ratios was futile and absurd (Kurunmäki and Miller 2013, 1108). The continued growth of ratio analysis was no doubt supported by the formation of the US Securities and Exchange Commission in 1934, in response to the stock market crash of 1929, and the increased availability of financial statements.

In the early 1930s, a number of studies sought to predict the likelihood of failure by focusing on individual ratios and comparing ratios of failed companies with those of successful firms. Several of these claimed to be able to identify predictors of failure some years prior to failure (Bellovary, Giacomino, and Akers 2007). Across the following decades, various bodies continued this quest for making failure quantifiable and probabilistic, including the Interstate Commerce Commission, the Internal Revenue Service, the Securities and Exchange Commission, the Board of Governors of the Federal Reserve System, and the Federal Deposit Insurance Corporation.
The calculative technology of ratio analysis had gradually become a more or less self-evident way of representing and calculating the condition of a company, both for external credit and for internal monitoring and control purposes.

In the 1960s, the aspiration to predict failure became an academic industry, initially based at the University of Chicago. Professor William H. Beaver matched a sample of failed firms with a sample of non-failed firms and studied their financial ratios for a period of up to five years prior to failure, concluding that three “non-liquid” ratios were the best predictors of failure (Beaver 1966). Others pursued a similar path, albeit using different sets of ratios and with differing weights attached to each, while multivariate statistical techniques and conditional probability models supplanted the more rudimentary models (Kurunmäki and Miller 2013, 1110). By the 1980s, the chain of calculations that linked financial statements, ratio analysis, and indexes of ratios was firmly established as a product that could be sold in both the academic marketplace and the world of financial services. The calculative infrastructure for economising both failing and failure was by this point firmly established, even if accurate predictions remained somewhat elusive.

These three steps – the forgiving of failure through insolvency legislation, the emergence and growth of credit rating agencies, and the development of a chain of calculations based on financial reports, financial ratios, and finally risk indexes – enabled both an economising of failure and the assembling of a failure regime for the corporate sphere that sought to identify failing and, in theory, forecast failure. It was the formation of this entire apparatus that enabled, in recent decades, the generalising of this idea of failure. These developments over a century and a half provided the preconditions for the attempts to economise failure in the public sphere, which we turn our attention to now.

Economising the Public Sphere: Assembling a Failure Regime for Hospitals in England

The economising of the public sphere is widely acknowledged as integral to the neoliberal reforms of recent decades. In contrast, much less attention has been paid to the more recent economising of failure, whether in healthcare, education, or social care. This despite the fact that an economised notion of failure, with its attendant possibility of exit, now dominates regulatory regimes, the entities they seek to regulate, and indeed much of the debate concerning the performance of public services. The contemporary language of failure goes hand in hand with a set of metrics and devices for calculating potential failure, for determining whether there are problems and, if so, how severe they are, setting out what might be done to address the problems identified, and if they are sufficiently severe, in pronouncing on the moment of failure itself. Exit, rather than voice, has become an option for dealing with severe decline and decay in the public sphere (Hirschman 1970). Insofar as public services are designed increasingly according to the rules of the market game, the entities providing them may now be allowed to fail according to the same rules, or so it has been suggested. At least in principle, bankruptcy law may as a result be equally applicable to the provision of healthcare and the corporate world. And the regulation of these very different domains can be circumscribed by the aspiration in a liberal society to ensure transparent and equitable arrangements for identifying failings and pronouncing on failure, yet without giving rise to a limitless expansion of the domain of regulatory intervention. To understand this most recent stage in the economising of the public sphere, we consider the attempts to assemble an economised failure regime for hospitals in England across the past two decades. These attempts demonstrate the multiple components that have to be assembled into a more or less functioning and stabilised ensemble before an economised failure regime can be put to work.
The year 2003, and the passing of the Health and Social Care Act in that year, can be taken as the
starting point for the attempts to assemble an economised failure regime for hospitals in England.
This entailed the creation of a new type of entity (called Foundation Trusts), a new definition
of failure, and a new calculative infrastructure which a newly formed regulator could then deploy. We
consider these three components of the emerging failure regime for National Health Service (NHS)
hospitals in England in the following sections. They provide an intriguing example of an attempt to
subject public hospitals to a financial discipline that turned out to be more rigorous than that applied
to banks. This contorted process began prior to the 2008 financial crash, and continued throughout
its aftermath and is still ongoing, although in a somewhat attenuated form as of the time of writing
due to the ongoing pandemic. Initially, it appeared to its protagonists as relatively straightforward,
insofar as the corporate model of failure, as represented by the Bankruptcy Act 1986, was considered
largely transposable to the public sphere, albeit with a bit of tinkering. But as the process unfolded,
and particularly in the light of the near collapse of the global banking sector, which coincided with
the eruption of one of the biggest scandals regarding care quality to beset the NHS, things became
increasingly complicated (Francis QC 2013).

We outline the protracted toing and froing of this process, for it tells us much about
marketising and economising, and the roles of changing ideas and instruments of failure in this.
To understand what is at stake here, one has to denaturalise failure, deprive it of its self-evidence,
which for some is a step too far. Surely, of all those things that have an undeniable facticity, failure
is right up there at the top of the list. Things have always failed, and sometimes financially, it
will be protested. This has nothing to do with the terms we use to describe such events, or the
calculations we perform when we reckon up the losses, so the protest continues. But nothing
could be further from the truth. Of all the things that are “made up,” that come into being hand
in hand with the ways in which they are named (and in this case calculated), failure is a perfect

To put this differently, and as noted above, failure is an archetypal variable ontology object
(Kurunmäki and Miller 2013, 1101). The actual moment of failure has none of the objectivity
and inevitability often attributed to it. That moment emerges within and through an assemblage
of calculative practices, expert claims and pronouncements, legal procedures, financial norms and
risk assessments, and much more. These allow a multitude of potentially conflicting interactions
among a wide variety of actors, aspirations, and instruments that illustrate only too clearly the
conditionality of performativity (see Butler 2010; see also Kurunmäki, Mennicken, and Miller
2016, 399). Or, as Hacking (2002, 106–7) puts it, the varying degrees of possibility which are
intrinsic to dynamic nominalism.

In the next section we examine the construction of a new type of entity (called Foundation
Trusts), and the co-construction of a new definition of failure. In the following section we
explore the construction of a new calculative infrastructure for a newly formed regulator. It
is through such multiple and interlinked interventions that the assembling of an economised
failure regime for hospitals in England was to be attempted.

A New Type of Entity, and a New Definition of Failure

On 30 April 2003, Alan Milburn – then Secretary of State for Health – delivered a speech to
the Social Market Foundation on the subject of healthcare provision. A new type of entity
was needed, Milburn argued. In place of hospitals that were exclusively line-managed NHS
organisations, a multiplicity of providers would be formed that would be governed by a simple
promise: the better the performance of the organisation, the greater freedom it will enjoy. A new
type of independent not-for-profit entity would be created, a sort of “third way” in healthcare
Economising Failure and Assembling a Failure Regime

(see for example Giddens 1998). Appealing to arguments on both the left and the right, Milburn spoke of the case for “new forms of organisation such as mutuals or public interest companies within rather than outside the public services and particularly the NHS” (Milburn 2002b).

This new type of entity was to be given a name. Those hospitals that were to be freed from day-to-day interference from central government, that were to be given local flexibility and freedom to improve services for patients, were to be called “Foundation Trusts.” Unlike NHS Trust hospitals, these new types of organisations for providing healthcare would be free-standing legal entities, no longer directed by the Secretary of State. They would occupy the middle ground within public services, located between state-run public services and shareholder-led private structures, yet still fully within the NHS. And, as central control over day-to-day management ceased, so should local community input be strengthened.

As free-standing entities, they would be held to account through the commissioning process rather than through day-to-day line management from central government. Commissioning had been introduced into the NHS in the early 1990s. It separated the purchasing of services from their delivery, creating an “internal market.” It was argued that making providers compete for resources would encourage greater efficiency, responsiveness, and innovation. Foundation Trust hospitals took this a step further, insofar as such entities would, for instance, have the freedom to retain proceeds from land sales to invest in new services for patients. They would have greater freedom to decide what they could afford to borrow, and they would be able to make their own decisions about future capital investment. They would also be given more flexibility with regard to pay, allowing “additional rewards for those staff who are contributing most” (Milburn 2002a).

However, this increased financial autonomy had as its corollary the possibility of failing financially, in much the same way as an entity in the private sector. On 20 November 2003, the Health and Social Care Act was passed. With this Act, the co-construction of a new entity and a new definition of failure was enacted. In an entire section headed “Failure,” the Act set out the procedures for dealing with NHS Foundation Trusts considered to be failing. This included provision for a newly formed regulator to intervene if a Foundation Trust were found to not be complying with its terms of authorisation. It also included provision for the regulator to remove any or all of the directors or members of the board of governors. And, importantly, it included provision for “voluntary arrangements” and “dissolution” consistent with the provisions set out in the Insolvency Act 1986. It also included provision for ensuring that the goods and services which the Foundation Trust had been providing continue to be provided, whether by the Trust itself or another body.

This aspiration to introduce a corporate model of failure within the NHS hospital sector was a big step beyond the many attempts to control costs in the 1950s and 1960s, and the subsequent and equally unsuccessful attempts to introduce delegated budgets in the 1980s. It brought with it profound tensions between the tripartite aspirations of local accountability and mutualism, an “exit” or insolvency model based on the corporate sector, and the retention of at least a residual form of central control in order to guarantee the continued provision of services. But what it did achieve was the first step in the assembling of an economised failure regime for hospitals in England. We turn in the next section to consider the subsequent step in the assembling of this failure regime: the devising of a new calculative infrastructure for the newly formed regulator.

A New Calculative Infrastructure for a New Regulator

If the first step in the assembling of a failure regime for hospitals in England was the construction of a new type of entity, and a new and economised idea of failure, the next step was the
construction of a calculative infrastructure which would allow the regulator to identify and act on both failing and failure. This again highlights the importance of emphasising the variable ontological status of both failure and the regimes established to identify and act on it. For the calculative infrastructures that enable such interventions are inherently relational. It is not a matter of devising metrics that will simply allow a given object to be recognised. Instead, it is a matter of devising a set of practices that form and fashion the idea of failure itself, in such a way that they can become the correlate of a historically specific set of practices for acting on it. And it is the relative stabilising of the relations among all the components of a failure regime that enables it to achieve a semblance of legitimacy, even if one or more of the components struggles to gain acceptance.

The newly formed Regulator was of course eager to deploy a set of metrics to assess the financial health of those Trusts that sought to become Foundation Trusts, as well as to monitor the performance of those that were authorised. Fortunately, McKinsey & Company were more than ready to help in this respect, and equally fortunately they had a set of metrics that were more or less ready “off the peg” from the corporate world (Kurunmäki and Miller 2008).

It was agreed at the outset that the Regulator’s role must focus on risk management. On the financial side, it was proposed that this would be based on key metrics such as liquidity, borrowing, and performance against financial projections provided during the application process. A balanced scorecard approach of sorts was suggested, to generate an annual risk rating that would determine the monitoring regime for the forthcoming year. A rating of one would indicate no cause for regulatory concern on any of the assessed components, and would result in biannual monitoring. A rating of five would suggest a high probability of a significant break of the terms of authorisation in the short term unless remedial action was taken. The stated overriding objective was to assess and mitigate potential risks to the delivery of Foundation Trusts’ obligations under their terms of authorisation. There remained, however, the thorny issue of what constituted a “significant failure” as set out in the 2003 Health and Social Care Act. There was also the dilemma that, while Foundation Trusts would enjoy considerable freedom, as the regulator pointed out “they can also become insolvent” (Monitor 2004).

The speed at which all of this was achieved was impressive, not least as the detailed specifications for authorisation still remained to be specified. That said, the metrics used were not exactly novel, bearing as they did a striking resemblance to those already in use in the corporate world (Dev 1974; Laitinen 1991; Moses and Liao 1987; Power 2007; Tamari 1964). Yet a working definition of the notion of “significant failure” still remained little more than an aspiration. While the principle of risk-based regulation was embedded rapidly in the new failure regime, the notion of failure itself proved more complex to operationalise in the healthcare context.

One consultation document followed another over the next few years, in the attempts to devise a failure regime based on the 1986 Insolvency Act, albeit one that allowed for the protection of essential NHS services and assets (Department of Health 2004a, para 2.9). These attempts ultimately floundered on the twin challenges of defining precisely what financial failure would mean in the hospital context and how it would be administered, and what the balance would be between financial failure, clinical failure, and governance failures. Many commentators voiced their concerns that financial failure would dominate, to the detriment of patients and/or the provision of mandatory services, but many also commented on the lack of clarity as to what would trigger intervention even on the grounds of financial failure alone. Other commentators remarked on the lack of an “administration” process, as per the 1986 Insolvency Act, which was described as a flexible and useful rescue tool, in the case of insolvency.
The objections raised proved overwhelming, and after five years of toing and froing the Department of Health concluded that “it is not appropriate to apply this quasi-commercial insolvency process to NHS Foundation Trusts or indeed to other state-owned providers” (Department of Health 2008, para 50). The language changed at this point from the economised notion of insolvency to the more nuanced terms “unsustainable provider” and “de-authorisation,” the latter meaning that a failing Foundation Trust would revert to being a NHS Trust, and a special administrator would be appointed to take control of the Trust. This shift in language and process was no doubt aided by the shocking events at the Mid Staffordshire NHS Foundation Trust, with the Healthcare Commission publishing a highly critical report in March 2009 which concluded that in its drive to acquire the status of Foundation Trust, it had “lost sight of its real priorities,” namely the quality of care it provided to patients, particularly those admitted as emergencies (Healthcare Commission 2009, 11). That report was followed by the announcement in June 2010 of a full public enquiry, which reported in February 2013. The preamble to the report spoke of the “appalling suffering of many patients,” and the consequences of “allowing a focus on reaching national access targets, achieving financial balance and seeking foundation trust status to be at the cost of delivering acceptable standards of care” (Francis QC 2013, 9).

These developments effectively put an end to the attempts in England to subject hospitals to a failure regime based primarily on corporate insolvency legislation. The Health Act 2009 enshrined this shift in legislation, by requiring the Regulator to consider, when aiming to identify failure, the health and safety of patients, the quality of services provided, the financial position of the trust, and the way in which it is being run (Department of Health 2009, 5). In place of a corporate model based largely or wholly on the notion of insolvency and exit, there was a significant broadening of what counted as failing and failure, and changes to the ways in which they could be identified and made operational in the context of healthcare.

Subsequent developments reinforced this broadening of what counts as failure. These included the removal in the 2012 Health Act of the “de-authorisation” option, as well as the provision in the 2014 Care Act for the Care Quality Commission to instruct the Regulator to appoint a special administrator where the care quality regulator observed a serious failure to provide services of sufficient quality; a provision that was only used twice between 2012 and 2014, and was itself not without critics (Murray, Imison, and Jabbal 2014, 24). That regime is now itself in flux, as the relationship between the various regulatory bodies involved is once again being reconfigured, along with their protocols.

That said, while the language of insolvency retreated into the background, the calculative infrastructure used to identify risks and financial failings remained in place, even if modified from time to time and placed alongside other non-financial metrics. Through these metrics, an economised notion of failure retained a significant presence, even if it did not have the dominant role envisaged at the outset.

**Conclusion and Further Lines of Inquiry**

We have argued in this chapter for greater attention on the part of sociologists to a particular aspect of economising, namely the economising of failure and associated failure regimes. More specifically, we have called for a focus on the emergence and assembling of specific failure regimes, and the co-construction and assembling of entities, ideas, and infrastructures. We have examined the economising of the economy, through the economising of failure for the corporate world across much of the nineteenth and the first half of the twentieth century. And we have examined how this notion of failure as an economic event in the corporate world has
travelled into the public sphere, and particularly the domain of hospital-based healthcare in England across the first two decades of the twenty-first century.

To address this phenomenon, we have identified four features of our approach to analysing the economising of failure. First, we have argued that failure has none of the objectivity or inevitability often attributed to it, whether in the corporate sphere or the public sphere. Second, we have suggested that failure be viewed as a variable ontology object, that researchers attend to the gradations in the stability of entities, agents, and infrastructures through which failing and failure are acted upon. Third, we have called for attention to the calculative infrastructures that operationalise the ideas of failing and failure, and enable them to be acted upon. For it is through the stabilising and intertwining of calculative instruments and ideas that failure regimes can be put to work. Fourth, we have emphasised the importance of distinguishing between failing and failure, for it is through this distinction, and the range of instruments that calibrate performance, that regulators adjudicate and determine the interventions that may be appropriate.

We argue that this conceptual toolkit helps us understand the multiple and co-constructed components in the assembling of a failure regime for hospitals in England. By multiplying or pluralising the dimensions of a failure regime in this way, we are able to better understand the outcomes of the processes of assembling and reassembling the various elements. This analysis helps us understand how an economised definition of failure can be preserved, together with the calculative infrastructure that makes it operable, even as the specific notion of insolvency recedes from view.

More generally, our analysis of the various attempts to economise the economy, and to economise the public sphere, provides insights into the limits of “actually existing” neoliberalism. Our analysis highlights the apparent tension between expanding the reach of market-based principles with the attendant rhetoric of local autonomy, while at the same time expanding the scope of central government to regulate and intervene, especially when things go wrong. This provides a timely reminder that marketisation and “decentralisation” do not signify a retreat of the state but a new and indirect way of governing individuals and entities (Béhar, Estèbe, and Epstein 1998; Donzelot 1984, 179–263). Initiatives such as those examined here, and which often go under the rubric of neoliberalism, cannot be understood by counterposing a non-interventionist to an interventionist state. They must be understood rather in terms of the novel ways in which they responsibilise those charged with day-to-day control of the entities they manage, and within parameters set elsewhere. The “autonomisation” of entities from direct control by the state allows, if anything, increasingly nuanced and detailed forms of intervention by the centre in a variety of ways.

The notion of exit, the possibility of failure, takes us to the heart of these issues. It takes us beyond the incessant measuring and comparing of performance. It highlights the difficulty of applying corporate models in an unfamiliar and even alien setting. It demonstrates the immense complexity of making a new entity that can be readily separated from the system in which it is embedded, and doing so while also making up the regulatory regime for that entity. It shows how attempts to transform passive patients into active consumers of healthcare are central to these changes. And it demonstrates the fundamental importance in all this of the calculative infrastructure of accounting and risk management, along with the associated and additional logics or discourses that have surrounded this initiative. Sociologists, we argue, should pay much greater attention empirically to the notions of failing and failure, and the assembling of the calculative infrastructures through which they are operationalised.
Notes

1 This chapter is based on work conducted as part of the programme of the Centre for Analysis of Risk and Regulation, and draws on Kurunmäki and Miller (2013); Kurunmäki, Mennicken, and Miller (2018; 2019). The authors wish to thank Mike Power for his encouragement with this project from the outset. We also gratefully acknowledge the financial support provided for this study by the Economic and Social Research Council (Grant Ref: ES/N018869/1) under the Open Research Area Scheme (Project Title: QUAD – Quantification, Administrative Capacity and Democracy). The QUAD project is an international project co-funded by the Agence Nationale de la Recherche (ANR, France), Deutsche Forschungsgemeinschaft (DFG, Germany), Economic and Social Research Council (ESRC, UK), and the Nederlands Organisatie voor Wetenschappelijk Onderzoek (NWO, Netherlands).


3 The US Congress approved both the Bankruptcy Act and the Reconstruction Act of 1867 (the first of four major provisions for readmitting former Confederate States) on the same day: 2 March 1867.

4 On the interaction between “lawmaking” at the national level and “norm making” at the global level, see Halliday and Carruthers (2007).


6 Pickering speaks similarly of “interactive stabilization” (see Pickering 1995).

7 On the aspirations for NHS Foundation Trusts, see Day and Klein (2005); see also Klein (2003; 2004).

8 Monitor Board meeting minutes, 20 April 2004, para 12.

9 King’s College NHS Trust’s response to the consultation document; Addenbrooke’s NHS Trust’s response to the consultation document; see also Department of Health (2004b). Here, attention was drawn specifically to S101 of Part IV of the 1986 Insolvency Act.

10 City of London Law Society’s response to consultation document; see also Department of Health (2004b).

11 On 9 June 2010, the Health Secretary Andrew Lansley announced a full public enquiry into the “commissioning, supervisory and regulatory bodies in the monitoring of Mid-Staffordshire NHS Foundation Trust,” to be chaired by Robert Francis QC and to report by March 2011 (Lansley 2010).

12 Put differently, “hybridization” has limits in some contexts. On the notion of hybridisation, see Kurunmäki (2004); Kurunmäki and Miller (2011); Miller, Kurunmäki, and O’Leary (2008).

13 On the issue of making an accounting entity, see Kurunmäki (1999).

References


Earling, Peter R. 1890. Whom to Trust: A Practical Treatise on Mercantile Credits. Chicago: Rand, McNally and Company.


https://doi.org/10.1136/bmj.328.7452.1332
Liisa Kurunmäki, Andrea Mennicken, and Peter Miller


