Open access • Journal Article • DOI:10.1080/19460171.2019.1594328

The incommensurables: the arduous art of making a regulatory indicator

Shirley Kempeneer, Wouter Van Dooren

Institutions: University of Antwerp

Published on: 02 Jul 2020 - Critical Policy Studies (Routledge)

Topics: Order (business)

The incommensurables : the arduous art of making a regulatory indicator

The Incommensurables: The arduous art of making a regulatory indicator

Shirley Kempeneer & Wouter Van Dooren

Abstract

The stress test of the European Central Bank has become one of the primary regulatory tools for the European banking system. In order to make such a regulatory indicator, different national banks need to be commensurated. They need to be made comparable according to a common metric. Despite a substantive literature, little empirical work has been done to further our understanding of the social and political processes through which these indicators are made. We use Actor-Network Theory (ANT) to enrich the existing literature with an in-depth account of how commensuration is negotiated. We find that despite a preference for commensuration, regulators allow 'incommensurable' categories to exist due to largely unrecognised regulatory benefits.

Keywords: Regulatory Indicators, Commensuration, Capture, Gaming, Actor-Network Theory (ANT), transnational knowledge controversies

How to make a regulatory indicator: the case of the stress-test

Mid-October 2013, 130 banks received a letter from the European Central Bank (ECB). Congratulations were in order; the banks were selected as 'significant financial institutions' in Europe. As such, they would fall under the ECB's Single Supervisory Mechanism as of November 2014. This meant that from now on they would be reporting to the ECB, instead of their own country. But before being admitted to the top league of European banking, the banks had to prove that they were healthy. To do so, the ECB conducted a 'Comprehensive Assessment', scrutinizing all 130 banks to test their financial resilience. A key part of the Comprehensive Assessment was a stress test. The ECB announced it would be the toughest stress test to ever be conducted in the European financial sector. 'There will be blood' a top official allegedly said[i].

The stress test assesses how well banks would cope during a three-year crisis scenario. More specifically, the test shows how much capital a bank still holds against its risk weighted assets after three years of crisis. The figure below (Table 1) shows an excerpt of the results (EBA, 2014). Even without a good understanding of how the indicator is made, it is possible to infer which banks are healthy (high percentages) and which are not (low percentages). The results made headlines globally. The reputational impact was significant, even before the results had any legal consequence. In the run-up to the stress testing exercise banks were seen to raise significant amounts of capital to increase their score, conscious that financial markets would be judging their results (Titcomb, 2014). In subsequent rounds of stress testing, the indicators were also formally linked to regulatory interventions, such as banks' Supervisory Review and Evaluation Process (SREP), which amongst others determines how much capital banks are legally required to hold.

Table 1. Stress test results

Name / capital ratio	2013 capital ratio (AQR adjusted)	Projected baseline capital ratio 2016	Projected adverse capital ratio 2016
Austria			
BAWAG PSK	14.3%	11.9%	8.5%
Erste Group Bank AG	10.0%	11.2%	7.6%
Belgium			

AXA Bank Europe SA	14.7%	12.7%	3.4%
Belfius Banque SA	13.5%	11.0%	7.3%

Regulatory indicators, such as the stress test, pervade transnational governance. Indicators and rankings play an important role in the way governmental and non-governmental organizations distribute attention, make decisions, and allocate scarce resources (Rottenburg & Merry, 2015). The OECD's Programme for International Student Assessment steers decision making in education policy (Gorur, 2016); the Human Development Index informs the United Nations Development Program (Davis, Kingsbury, & Merry, 2012); and the EU's Open Method of Coordination helps policymakers monitor and measure progress in various policy domains across member states (Marlier & Atkinson, 2010). We define an indicator, following Davis et al. (2012: 75), as:

> A named collection of rank-ordered data that purports to represent the past or projected performance of different units. The data are generated through a process that simplifies raw data about a complex social phenomenon. The data, in this simplified and processed form, are capable of being used to compare particular units of analysis (such as countries or institutions or corporations), synchronically or over time, and to evaluate their performance by reference to one or more standards.

We emphasise the regulatory character of such indicators, as they can steer behaviour even without legal enforcement. The reputational pressure that indicators exert with their potential to rank performances, is often effective in in securing compliance or behavioural change (Grabosky & Braithwaite, 1986; Tervonen-Gonçalves, 2012; van Ostaijen & Scholten, 2017).

Several reasons present themselves to study this process of how transnational indicators are made. First of all, commensuration, the need to homogenize across contexts, reaches its pinnacle in the European context (Bruno, Jacquot, & Mandin, 2006; Mügge, 2016). Despite the diversity in national contexts, Member States, and their policy issues, are intricately connected in many ways, calling for an overarching European regulatory system (and legislation) (Jurgen Habermas & Derrida, 2003; Kohler-Koch, 1996). This became

especially clear in the financial crisis where the interconnectedness of the banking system caused risk to spill over national borders and contaminate the entire European sphere (De Bruyckere et al., 2013). The complexity of transnational policy issues simultaneously enhances both the need for clear-cut indicators to regulate, as well as the difficulty to create such indicators across varying contexts. As such, efforts towards, as well as struggles with commensuration will likely play a pivotal role. Discussions on transnational commensuration and standardization tie in with Barry's (Barry, 2012)work on transnational knowledge controversies. One of the critical difficulties in governing transnational issues, such as financial policy, is a lack of transnational consensus on matters of fact, or how evidence should be interpreted.

Secondly, Welsh (2017) calls for a more critical and political analysis of rankings. Likewise, Mügge (2016) explicitly calls for political scientists in particular to pay more attention to the forces that determine indicators' design. Since the 1960s, political scientists have used theories of regulatory 'capture' to explain regulatory outcomes and designs. It conveys a sense of illegitimate expropriation, performed by one powerful group over others (Baxter, 2011). A substantial body of literature studies the privileged interactions between industry and public authorities (Bunea, 2013; Hanegraaff, Beyers, & De Bruycker, 2016; Klüver, 2013; Lowery, 2013). When indicators paint a predominantly positive picture of regulated entities, this is often ascribed to regulators designing a less critical performance measure (Woll, 2014). In the case of the ECB's banking stress test, such accusations were echoed in public debate as well. The crisis scenario that the ECB banks were subjected to was deemed far less severe than that of the Federal Reserve or the Bank of England, leading to misleadingly positive results for the European Banking Sector (Cecchetti & Schoenholtz, 2016). Beyond 'capture', public administration scholars criticize indicators for their susceptibility to 'gaming' (Bevan & Hood, 2006). Aware of which behaviour would be measured by the indicator, organisations or countries are seen to manipulate performance outcomes to their advantage, painting an overly optimistic picture. However, it is yet to be studied how these mechanisms come into play in designing regulatory indicators.

In what follows we build on literature in quantification (Desrosières, 1998; Espeland & Sauder, 2007; Espeland & Stevens, 1998; Hacking, 1990; Peeters, Verschraegen, & Debels, 2014; Porter, 1995) to improve our understanding of how indicators are designed. This is increasingly important, because they affect our understanding of the world in subtle and

often unrecognised ways (Rottenburg & Merry, 2015). As such, before we can use these numbers in policy and public debate, we need to understand precisely where they come from. We conduct in-depth interviews with stakeholders in risk departments in Belgian banks, consultants, and the ECB, to better understand how the stress test is made, and to which consequence.

We find that although these design choices are politically motivated, there is more to the story than mere gaming or regulatory capture. Introducing incommensurability can be seen as a manifestation of a critical epistemological attitude that is objected to a rationally calculable reality. In the concluding section, we reflect on what these findings mean for broader academic and public debate, relating it to the Frankfurt School's critique of instrumental reason (McCarthy, 1990) and more specifically to Habermas' notion of communicative rationality as an alternative to instrumental rationality (Habermas, 1990).

Commensuration and the production of regulatory indicators

Commensurating systems, means taking diverse qualitative systems and homogenizing them on a common metric, facilitating comparison (Espeland & Stevens, 1998). Literature on commensuration pays particular attention to the socio-political forces that lie behind this standardization process. Regulatory indicators used in transnational governance, commensurate systems over national contexts. They assume that is it possible and desirable to compare complex systems across countries according to uniform measures. In our case, the ECB stress test aims to make a standardized comparison of banks' risk across Europe. Commensuration literature stresses that things 'are' not comparable, but that they need to be made comparable. It thus proves interesting to pay special attention to how this commensuration process feeds into transnational indicator design, and how indicators make diverse systems comparable.

Commensuration is a social process. It begins with the idea that it is meaningful to compare a set of things. For example when estimating the costs of a large infrastructure project, we now believe that it is important to take into account the potential loss of natural resources, and other costs to the environment (Vickerman, 2007). To do so, we need to find a way to compare these costs and benefits. This is often done by putting a price on the loss

of land, or the quality of air, that we then can compare to the cost of traffic congestion or employment. However, practically finding ways to value and compare diverse inputs is no small feat (Patterson, 1998). As such, commensurability is not only a social construction, but a social accomplishment that requires substantial efforts. Entire agencies, industries and even disciplines are dedicated to finding ways to compare (the value or performance of) different systems (Jasanoff, 1986). In our case, it was only after the crisis that the urge arose to create a standardized pan-European comparison of banks, before this, each National Competent Authority had its own method of assessing banks' health. These different national strategies were then also tailored to the local context and banks' business models. Finding a way to compare these banks on a Level Playing Field across Europe was thus a challenging task for the stress test. Especially when it comes to big, systemically important, banks. Standardised reporting requirements can make it seem as if banks are easily compared across contexts. However, when you take a closer look, the financial products that these big banks hold are so disparate and complex, making it very difficult to calculate and quantify their value and risk. How much an asset in a big bank is worth, or what its risk is, can be an as vexing question as how much a human life is worth (at least in terms of all the different parameters that can be taken into account[ii]). A wide range of qualitative properties of the asset need to be transformed into quantities, and this can happen in a variety of ways, according to various assumptions, theories and models.

Just as things can be seen as commensurable, they can also be seen as 'incommensurable', or undesirable to compare. For example, Ackerman and Heinzerling (2005) argue that it is morally wrong to compare the value of human lives, health, or the environment to economic gains, especially not by reducing them to 'cold dollars'. These things are seen to be 'priceless', and thus incommensurable. Incommensurability is as much a social construction as commensurability, and also requires work (Espeland & Stevens, 1998). Incommensurability claims are often supported by moral arguments, that for example, human life should be valued above anything else, or the environment should be protected at any cost (Ackerman & Heinzerling, 2004). In the case of the stress test, banks argue that it is not fair to compare one portfolio to another at face-value. They claim that their assets are simply too different in nature to be assessed according to the same standards.

A more substantive critique comes from the Frankfurt School, who condemn the wider economic, political and social effects of commensuration (for a good overview see Smulewicz-

Zucker, 2017). In the Dialectic of Enlightenment for instance, Horkheimer and Adorno (1947) discuss the standardising effects of mass media, akin to a 'culture industry' producing uncritical identical individuals, locking in power relations, and reproducing dominant discourses. In a similar fashion, indicators (and other technologies) can act as a standardising instrument for control and domination (Marcuse, 1941). Treating banks as incommensurable would then be the only way to allow critical rational debate to triumph over manufactured information. Here incommensurability is not defended from a moral point of view, but from an emancipatory one.

Regulators are typically in favour of commensuration (Gorur, 2016; Porter, 1995; Scott, 1998). The illegibility of local contexts is an administrative headache for regulators. Without comparable units of measurement, it proves almost impossible to monitor, compare, or regulate performance in various policy domains. Regulators need tools like standardised indicators to understand and manage the large and complex reality. On the other hand, regulated sectors typically fight these commensuration efforts. They see their unique qualities stripped away and do not feel accurately represented by these standardised measures. They often have their own distinct view and contextual interpretation of the categories that are taken into account (Peeters et al., 2014)

In what follows we analyse this process of commensuration empirically, with the case of the ECB's banking stress test. To do so we use a framework borrowed from Actor-Network Theory (ANT), Callon's (1984) 'Sociology of Translations'. We elaborate on this approach in the methodological section.

Methodology

Our case study is the stress test of the European Central Bank (ECB). The stress test has earned its place in the centre of financial regulation. As it is a relatively new indicator, the design is still in full development. This makes it an ideal case to understand how and why design choices are made. We conducted 32 in-depth narrative interviews with risk experts in all four of the Belgian banks included in the ECB stress test, the consulting firms assisting the banks, as well as with ECB officials in Frankfurt. We contacted the Chief Risk Officer in each bank and snowballed from there, we also asked banks to refer us to consultants they worked with. As risk teams in banks were fairly small, we were able to speak with the most relevant people on

several occasions. At the ECB, we contacted respondents from both the macro- and micro-prudential fields, who were working on stress-testing, only one respondent that we contacted refused an interview. We recorded all but two interviews and transcribed them verbatim[iii]. We used a semi-structured topic guide, based on Callon's (1984) framework of translations. We began with open ended coding close to the text and then analysed and structured our data according to key concepts.

As a research heuristic we use Actor-Network Theory (ANT), and more specifically Callon's (1984) 'sociology of translations'. A basic assumption of ANT is that everything we see in the world is built up from a set of relationships between people and things, a so called 'Actor-Network' (Latour, 1999; Law, 1999). Callon uses the notion of 'translation' to explain how an Actor-Network, comes to be represented by a single thing, in our case an indicator of health. The linguistic metaphor of translation emphasizes the manner in which interests, goals, or desires are represented, simplified, and transformed in the production and mobilization of artifacts.

The logic of translations is helpful in studying processes of commensuration. Callon (1984) distinguishes four key moments of translation: problematisation, *interessement*, enrolment, and mobilisation. To untangle the concept of (in)commensurability, it is helpful to unpack it according to the various stages. For instance, during problematisation we can look into why and for whom commensuration is necessary, interessement and enrolment give us an understanding of how commensuration is negotiated in practice, and mobilisation can help us understand how commensuration ties into wider societal processes. This analytical framework thus gives us a more multi-facetted and critical understanding of commensuration and claims of incommensurability.

How to make a policy indicator: commensuration in action

We will structure this empirical section using Callon's four key moments of 'translation': problematisation, interessement, enrolment, and mobilisation (Callon, 1984). Looking at commensuration as a process of translations, gives us a more in-depth understanding of the role of incommensurables. The inability to commensurate is often written off as a failure in a process, rather than a deliberate action. Throughout the different translations, we gain a better understanding of what drives (in)commensurability and how.

Problematisation: should banks be commensurated?

In the first step of the assembly of an actor-network, a given actor analyses a situation and provides a specific problem definition. An important requirement is that this problem definition, and the subsequent proposed solution, rings true to other actors: What is the problem that needs to be solved?

In the case of the stress test, the 2008 financial crisis made clear that something was wrong in financial regulation. But what? National supervisors, along with the rest of the world, failed to see the crisis coming. A risk director in a bank noted: 'Before the crisis a lot of banks used quantitative models (…) but then people saw that a lot of banks failed and had issues during the crisis even if the models said everything was ok.' This illustrates how old indicators of banks' health showed significant shortcomings during the crisis. Moreover, there was a rising distrust towards national regulators. They were said to be 'captured' by the banks, no longer safeguarding the public interest but preoccupied with national interests and 'their banks' looking good. Banking supervision was concentrated at the national level before the crisis, because national authorities stood closer to banks and were seen as better equipped to understand the complex legal and socio-political context they operated in. A risk expert noted:

> You cannot take an asset in this bank, and just compare it to an asset that might look the same in another bank. You have to understand what's underneath this asset. For example, credit quality, we have clients with good savings accounts. That's typically Belgian actually. This is going to affect the PD [Probability of Default] and the LGD [Loss Given Default] and such.

This shows that comparing banks is not so straightforward. Before the crisis banks across Europe were treated as 'incommensurable' from a regulatory perspective: Banks were seen as too different in terms of business models, activities and portfolios, to be compared in a uniform way. Although there were general guidelines, banking supervision differed across countries. This changed after the crisis. The crisis revealed how interconnected banks in Europe were, and that despite their differences, it would be meaningful to compare the systematically important institutions. The problem shifted from a localised national issue, to an integrated European issue. This led to the establishment of the Single Supervisory Mechanism (SSM) in 2014[iv], that transferred supervisory power to the ECB. And as such, finding a way to commensurate these different, yet interconnected European banks became a key priority for the stress test.

Interessement and enrolment: To commensurate, or not to commensurate?

The next translations are interessement and enrolment. In interessement actors seek to lock other actors into specific roles. Actors, in ANT can be both people and things. To make this interessement successful, Callon speaks of enrolment: these are the negotiations that accompany the interessements and enable them to succeed (Callon, 1984). Think of interessement as writing up a script, and enrolment as negotiating everyone's part. In this section, we discuss the different actants that were 'interested' and 'enrolled' into the stress test, and the role they played in commensurating banks. In the stress test the process of

interessement and enrolment embody how the ECB wrote up the methodology (or script). By looking at which rules (or roles) were instated and how, we can understand how they play into processes of commensuration and claims of incommensurability.

Commensuration at any cost?

The starting point for the stress test is banks' balance sheet. Balance sheets give an overview of banks' assets and liabilities, these can be categorised in different sub-groups with specific characteristics. The stress test projects the impact of a three-year crisis scenario on banks' balance sheet. In reality, banks would make changes to the items on their balance sheet during a crisis, such as selling off bad assets. However, in the name of the LPF, banks' balance sheets were kept static over the three-year scenario. Some banks might make bigger changes than others, and it would be hard to compare the end results. The management decisions that a bank would make were not 'translated' into the stress-testing exercise, because they were impossible to predict and/or standardise. Keeping a static balance sheet was thus the only way to keep banks commensurable. This came at a cost. As we saw during the financial crisis, banks are intricately interrelated. The only way to map these interrelations, and understand their effects, is through a dynamic balance sheet. A respondent at the ECB explained:

> Say a bank's balance sheet is hit, maybe they're going to be selling off corporations, so that's going to spill over to other countries, maybe, who knows. But this is what we would like to know. But without a dynamic balance sheet, we cannot know.

A static balance sheet makes it possible to make a fair comparison between banks, but it stands in the way of predicting how a crisis would affect the banking system in its entirety. Here we see clearly that commensuration, and comparability, limits the questions that indicators can answer. An indicator design that focusses on comparing how banks react to a crisis scenario, will not be able to thoroughly answer questions about how these banks, and the banking system in Europe, would react to the scenario. The choice for a static balance sheet is a pragmatic one, that has little to do with steering results in any direction. Respondents in banks and at the ECB agreed that a dynamic balance sheet would be more

adept at gauging a bank's true risk. However, respondents understood the choice to keep the balance sheet static to facilitate fair comparisons.

A next important design choice is the adverse scenario that the banks have to face. The stress test was designed with a common scenario, meaning all banks would be subjected to the same macro-economic turbulences. This was – again - motivated by the LPF. The ECB wanted to compare banks under the same scenario to eventually assess which banks performed better or worse. At first sight, a common scenario would facilitate the commensuration of banks. However, a respondent at the ECB criticized this choice:

> What you want is a similar degree of pressure applied to all banks. But that does not happen with this single scenario. Banks are complex institutions, and each bank is sensitive to different things – it's like putting the same weight on different bridges.

So, on the one hand, it would seem fair to expose the sample of banks to the same scenario, but on the other it would also be fair that each bank would be exposed to the same amount of stress. This hints at the complexity of the questions that commensuration raises: Is it possible or even helpful to compare how different banks react to a crisis? The result of the commensuration might have less to do with the actual health of the bank, and more with inadvertent sensitivities that a bank might have to a specific scenario, or even sheer luck.

Another important design feature is the 'common methodology'. The common methodology describes in detail how banks are supposed to calculate and report the impact of the scenario on their balance sheet. This document was a key in the commensuration process. It forced banks to enter their balance sheet data according to standardized reporting standards and implement common definitions and assumptions to calculate asset values and risks. All banks filled in identical excel sheets with hundreds of data points, in identical columns and rows, that could easily be compared. Moreover, caps and floors were added to data points to further restrict large divergences in results. We asked where these caps and floors came from, and if they favoured any particular bank. A risk expert explained:

> No, I don't think they favour anyone. They really are just put in place to keep the results conservative, to make sure the results of the banks are not too far apart. They don't make any economic sense either, in my opinion. They just 'assume' for

> all banks that for example some results cannot be positive, or that you have to calculate something according to [a set of rules]. And in our case, we have some exceptions to these rules, and we can explain this. For some assets we provide an extra insurance or so. But we're not allowed to take that into account. We have to calculate everything the same.

We pushed the respondents in banks on their ability to influence the design in their favour, the response was:

> Well we talked with [an association of banks] but the thing is, it's hard to align interests, something that would be good for us is, is not necessarily good for another bank. So, it's not like you can press on these issues together. Even within Belgium. We did make the same arguments on some points, I called [a risk expert from another Belgian bank] to ask how are you going to interpret this rule, so we took a similar approach there. But these opportunities are limited.

This shows that 'capture' becomes more difficult when the interests of a group are not aligned, and the group sees its interests as 'incommensurable'.

The common methodology with the caps and floors seems to point in the direction of more commensuration efforts by the ECB. However, the fact that there is a common methodology, and banks are allowed to calculate the exercise themselves is quite remarkable. The ECB could have just conducted the stress-test themselves and left banks out of it completely. This is called a top-down stress test, where the central bank uses their own data or data delivered by banks, to calculate the impact of the scenario on banks. Using top-down models would take away any leeway banks would have to manipulate or game the exercise. Given the importance of the LPF, this would not have been a strange decision. Yet, the ECB opted for a bottom-up stress test, where banks were given some freedom to calculate the impact of the scenario themselves (albeit according to strict methodological guidelines). The question is: why?

Claiming incommensurability

We can be fairly sure that the reason is not technical infeasibility. The ECB has already developed top-down models to calculate the impact of the scenario on a bank. They just choose to solely use them for benchmarking purposes. A respondent at the ECB explained:

> The supervisors use the top-down results as a benchmark to judge the banks' results. And then there is a back and forth process with the banks. They can try and explain why their results are different from what we expect.

In this, the design allows some room for incommensurability. Banks are given the room to argue that their assets are incommensurable with seemingly similar assets in other banks. They can explain why their risk should be calculated differently than risk in other banks[v].

During interviews three reasons came up for designing the exercise bottom-up. The first had to do with banks being able to give the best representation of their risk themselves. A respondent at the ECB noted:

> Well I think banks should manage their own risk, I don't think supervisors need to do this. We really, and this is important, we want to foster the development of banks' own risk management capabilities. It is normal that banks should have a much better understanding of their risk. That is a big argument for a bottom-up stress-test. (…) From banks, it is not just a wilful act to spin figures, it's important to have a close discussion with institutions.

A risk expert in a bank added:

> You can't just take an asset class and treat it the same in different countries. When you discuss this in international committees sometimes it looks as if each country is just defending their interest. But for example, mortgage lending, in some countries you can give back the keys, in others you still have to pay, you are still liable. So historically the defaults on the loans are very low. So, there is an

> argument to be made that these banks do have safe portfolios and you should not apply the same risk weight globally.

This shows that supervisors at the ECB and banks agree that banks' assets should be treated as 'incommensurables'. Here, an accurate understanding of the risk of a bank's asset, is valued higher than the commensurability of the asset. It is agreed on that it is not always desirable to compare the risk of even seemingly similar assets. To be sure, we've seen that accuracy has been sacrificed before in the name of commensuration. So, this still leaves us with the question of why the incommensurability is granted this time. Technically, it would be possible for regulators to run the whole exercise themselves. Yet, banks were allowed to use their own internal models. This can be better understood by the second reason that was given for granting incommensurability.

The second explanation for the bottom-up stress-test had to do with accountability and responsibility. Commensuration can shift responsibility away from regulated systems, onto regulators. A respondent at the ECB stated:

> The team is great, but to do a top-down for all the banks in the comprehensive assessment, that's about 130 banks [in 2016]. I'm not sure that will happen. The inherent danger of using the same models for 130 institutions, that's a risk in itself." A risk expert in a bank elaborated: "The regulators can do this top-down exercise. But I feel that supervisors are a bit apprehensive that if they do everything themselves, calculate it, publish it, then they are accountable. If a bank gets a good score but gets in trouble the year after, well the supervisor will be blamed fully.

This suggests that supervisors do not want to be completely responsible for assessing what will happen to banks' assets in a stress-scenario. Especially, because supervisors see a danger

in banks using these top-down models, instead of developing their own models. A respondent at the ECB worded this carefully:

> We never give all the information about our {top-down} models. Just enough to understand the model, but not enough to replicate it. We don't want banks to just take the models and use them. Then we would lose the bank-specific models, which are obviously valuable. We do not want a mono-risk culture. There is a big top-down, bottom-up discussion globally. It is good if banks use our models as inspiration, but to replicate, no. you want to keep some uncertainty, because you have model uncertainty. You cannot say 'this is the one model', you can't put all the eggs in one basket. This should never become some unilateral guidance to banks.

This shows that even when commensuration is technically feasible, regulators choose to treat banks as 'incommensurables'. These reasons tie into Barry's (2012) notion of transnational knowledge controversies. Regulators cannot seem to choose one standard model to assess how banks' assets would perform under stress. There is no pan-European consensus regarding how much risk assets hold and how they would be affected under crisis situations. A consultant explained the trend over the past years:

> The idea was at first to apply standard risk weights to standard asset categories. But as banks and their financial instruments became more complex, these standard risk models did not reflect the true risk anymore. A few years ago, there was a clear direction to more sophisticated models, giving more flexibility to the bank to develop internal models that would really make the bank able to simulate the exact risk of their bank and business model. But we've seen that banks are a bit using or playing with these models to go around the rules. So, we are seeing a clear trend towards more standardized models, and more simple models. Because the big problem with all these

> complex internal models is that you get results from the different banks and you cannot compare easily. When you see that for the same exposure, banks have different RWA [Risk Weighted Assets], that's not normal.

This shows that regulators are struggling with building transnational knowledge, and a transnational consensus on how to gauge the risk of banks' assets. In order to commensurate banks across Europe, transnational guidelines need to be established regarding how risk should be perceived and calculated. However, as each bank is developing its own complex financial products, it becomes even more challenging to create a homogenized understanding and calculation of risk. This ties in with the third reason for the bottom-up stress test: the quest for evermore data. A respondent in a bank elaborated:

> Each time, the data requirements are even more granular. We're at what, 395 000 data points now? It's clear that the ECB wants this data. We saw that last time we were in Frankfurt, the apparatus that they've designed there is so sophisticated. They take this data and they do all these different things with it. I'm not sure what they're doing exactly, but they're doing a lot of calculations.

This shows that regulators are interested in gaining as much information as possible about the different banks. If they would conduct a standardized top-down stress test themselves they would lose this valuable information.

In conclusion, this section shows us that on the one hand regulators go through a lot of effort to commensurate systems, but on the other that commensuration is no regulatory panacea. To be able to compare banks on a level playing field, concessions are made in terms of accuracy. However, commensuration is not pushed at any cost. Despite the possibility of designing a completely standardised top-down stress test, regulators consciously opt for a bottom-up exercise, treating banks assets as incommensurable. By allowing banks to use their

own models, regulators create room for potential gaming efforts, where banks could try to make their portfolio look as good as possible. Designing a 'gameable' indicator may seem like a concession to industry interests, showing weakness from regulators side. However, this analysis shows that regulators consciously treat banks as incommensurable because of the regulatory benefits this entails, such as shared accountability, information access, and improved internal risk management capacity. Regulators are apprehensive in establishing transnational guidelines on how risk should be interpreted and calculated, because there are still controversies regarding the right way to calculate and measure financial risks. As such, if regulators were to establish a transnational norm, they would be fully responsible for the consequences in case the norm would prove erroneous. Moreover, they believe in the benefits that come from leaving the debate open, and continuously reconsidering how risk should be understood and calculated. This can again be related to a more critical epistemological stance, where regulators choose to treat banks as incommensurables from an emancipatory point of view.

Mobilisation: designing for results

The final step in Callon's framework is mobilisation: The actor-network starts to operate. This is the moment of truth, the final translation. Will the story hold, will all actors stick to their roles? The alliances made and consensuses agreed upon can be contested at any moment. Translation can become treason.

It is important to note that the result of the stress test needs to be able to 'hold' in the real world. It needs to be able to interact with other actors. As such, it's shaped by these possibilities for interaction. The stress-test is not created in a vacuum. In the mobilisation stage, it needs to be translated into the real world, and work there. A respondent at the ECB

summarised this well with the sentence: 'you're stuck with what you can afford'. The scores on the stress test become a new fact and other actors will interact with this fact, they will make decisions based on this fact. So, you pre-emptively need to take into consideration what decisions these actors might make, and which decisions you can 'afford'. The respondent elaborated with a concrete example of how financial markets would react to the stress test, and which reactions the US could afford, and which reactions Europe could afford, he finished off stating:

> So, it's obvious that the stress test results were mild. But this is not because the people who do these things are incompetent, or captured by banks, or are weak intellectually or whatever. There is also the dimension that it would have been irresponsible to come out with a cap request of 100 billion in such a situation [a conservative fiscal stance in Europe]. You're stuck with what you can afford.

A risk expert in a bank agreed: 'It's politically motivated, an exercise like this. They know exactly how severe the stress will be on the banks. They developed it like that. Those hundreds of pages of rules, they know approximately what the outcome will be. They know perfectly, with the exercise they drafted now, they know that is the message they want to be spreading. And they do so in everyone's best interest.' As such, the message was clear from the beginning: the stress test would have to say banks overall were healthy. Indicators have important performative functions, they can become real in their consequences. If the ECB would come out with the message that banks were unhealthy, financial markets would have reacted this, only aggravating the situation. As such, the design of an indicator does pre-emptively need to take this performativity into consideration, especially when making results public.

We confronted respondents with this. If the results were decided on from the beginning, why did we need such an elaborate exercise? Here the notion of credibility came up a lot. A consultant phrased this well: 'The ECB needed to show that they really knew what was going on in the banks. They needed to take a deep dive. And they needed to set common ground rules for all the banks. We saw this in the early CEBS exercise. It was worth nothing, and nobody believed it. Banks were still all doing whatever they wanted with all the national discretions. It said the banks were all fine, and then after publication we saw banks failing. That's why the ECB needed this elaborate rule book for banks to follow.'. This shows that in order to be a credible indicator, commensuration plays an important role. Supervisors cannot just produce a list of numbers, and hope that the wider public will believe them. These numbers need to be made credible through processes of commensuration.

How a regulatory indicator is made: lessons learned

The production and use of policy indicators in global governance is increasing rapidly. Indicators have the unique power to simplify policy issues and rank performances according to a simple numerical scale. As such, they have important regulatory effects, even when they lack any legal mandate. In order to make a regulatory indicator, different (national) systems need to be commensurated, i.e. be made comparable according to a common metric. Despite a burgeoning literature, little empirical work has been done to further our understanding of the social and political processes through which these indicators are made, and units are commensurated (Davis et al., 2012; Huault & Rainelli-Weiss, 2011; Peeters & Verschraegen, 3013). As such, we conducted a case-study of the European Central Bank (ECB) stress test, interviewing people at the ECB, consulting firms and Belgian banks. Where it is often assumed that regulators have a blind preference for commensuration, and standardisation at any cost (Gorur, 2016; Scott, 1998), we show how regulators allow 'incommensurable' categories to exist due to largely unrecognised regulatory benefits.

This work draws on and contributes to the analytic tradition of quantification and governance by numbers, as developed by Desrosières (1998), Power (2003), and Rose (1991), and more closely to literature on commensuration (Espeland & Stevens, 1998; Kolk, Levy, & Pinkse, 2008). At the same time, it is located in relation to STS accounts, such as ANT and Callons' sociology of translations (Bijker & Law, 1992; Callon, 1984; Latour, 1987; Latour & Woolgar, 1979). What we add to this literature is an in-depth empirical understanding of how commensuration is done in practice, and why, giving us a better understanding of where the numbers we use come from. The bulk of the literature so far focusses on either how commensuration can be achieved technically, or why it is problematic and should be avoided. In this paper we use Callon's (1984) framework of translations to unpack commensuration as a multi-faceted process. This gives us a more detailed notion of the motives underlying the commensuration process, as well as its benefits and drawbacks.

For starters, The results of the ECB stress test were seen as 'mild' and biased in favour of banks (Cecchetti & Schoenholtz, 2016). In both public and academic debate, regulation is seen to be routinely 'captured' and manipulated to serve the interest of regulated entities (Dal Bo, 2006; Etzioni, 2009). However, we find that regulatory capture proves to be difficult where interests of the regulated sector lie so far apart, and the sector sees itself as incommensurable. Additionally, we find that although regulators could have made a more 'game-proof' indicator, their choice to abstain from doing so did not result from industry pressure but was a conscious choice. This adds an important nuance to theories that use regulatory capture to explain regulatory outcomes and designs (such as Bunea, 2013; Hanegraaff et al., 2016; Klüver, 2013; Lowery, 2013; Woll, 2009).

Secondly, we challenge the idea that some things 'are' more commensurable than others. Banks across Europe are not automatically comparable, they need to be made so. This is often overlooked in policy fields like the financial sector, where quantification is taken for granted. Our findings emphasise that commensurating systems is a socio-political construction that requires a lot of effort. In this vein, we argue claims of incommensurables can be made, and should be heard, in all policy fields. Allowing political actors to make claims of incommensurability, is a way to emancipate them and give them a critical voice. As such

this argument ties into critiques of instrumental reason (for instance, see Smulewicz-Zucker, 2017).

Moreover, we find that commensuration can contribute to credibility. A large part of the credibility of the stress test results from the extensive commensuration efforts, as proof that the ECB is assessing European banks on a Level Playing Field, subjecting them to extensive uniform rules. As a key policy indicator, it was imperative that the stress test was taken seriously by market participants. If the ECB would have published a list of banks' health that was not taken seriously by the wider public, this could have aggravated the crisis.

Finally, however, we find that banks are still treated as incommensurable to some extent. The ECB does not assess banks' assets according to standardised models. Rather, banks are allowed to use internal models for calculation. This may seem as if regulators are giving banks free play to game the indicator, and make their results look as good as possible. However, this choice for incommensurability is informed by other regulatory benefits, such as shared accountability, information access, and improved internal risk management capacity. Although literature usually promotes the design of game-proof indicators (Bevan & Hood, 2006; Hood & Peters, 2004; Politt & Talbot, 2004; Smith, 1995), we argue that allowing room for gaming can have regulatory benefits that have largely gone unrecognised.

This ties into Barry's (2012) notion of knowledge controversies in transnational governance. Regulators shy away from establishing an extensive transnational consensus on how risk should be understood and calculated. On the one hand, they are apprehensive of bearing the responsibility of establishing a standard model of measuring risk. On the other, they seem to believe that knowledge controversies, to a certain extent, allow for innovation, as they maintain a continuous reconsideration of how to understand and measure risk. This goes to show that technological instruments, such as the stress test, play a critical part in

developing regulatory spaces (as also argued in Barry, 2001). This non-instrumentalised conception of knowledge is also reminiscent of Habermas' (1984) notion of communicative rationality as an alternative to instrumental rationality. The former focusses more on increasing understanding through open communication, while the latter is more strategic and results-oriented. Processes of commensuration tie in closely with instrumental rationality; as they both aim to manipulate the world in order to control it. Following this, indicators can act as a standardising instrument for control and domination (Marcuse, 1941). Treating banks as incommensurables then reconceives knowledge making as an ongoing exchange among critical equals, rather than a fixed outcome of a so-called rational process imposed by dominant actors. Allowing incommensurability establishes discursive conditions that offer a more critical and understanding-oriented space for the regulatory exchange. Rather than trying to merely measure and control risk by imposing 'rational' knowledge, regulators here make a more critical attempt to understand risk. Recognising banks as incommensurables thus marks a noticeable shift from an instrumental to a more critical epistemology in financial regulation.

[i] According to respondents, Daniele Nouy, the chair of the Supervisory Board at the ECB, made this statement while announcing the stress test.

[ii] And when you hear experts talking about this it often seems as if there is as much at stake.

[iii] Two respondents at the ECB did not want our conversation to be recorded, but we were allowed to take notes, which we subjected to the same coding as the transcribed interviews.

[iv] For more on the SSM see the website of the ECB: https://www.bankingsupervision.europa.eu/about/thessm/html/index.en.html

[v] To be sure, the ECB is quite strict when it comes to granting exceptions. In discussions the ECB does often still have the upper hand and can force banks to comply with the results they projected.

References

Ackerman, F., & Heinzerling, L. (2004). *Priceless: On Knowing the Price of Everything and the Value of Nothing*. New York: The New Press. https://doi.org/10.1136/bmj.330.7499.1091

Barry, A. (2001). *Political Machines : Governing a Technological Society*. London: Bloomsbury Academic. https://doi.org/10.5040/9781474213110

Barry, A. (2012). Political situations: knowledge controversies in transnational governance. *Critical Policy Studies*, *6*(3), 324–336. https://doi.org/10.1080/19460171.2012.699234

Baxter, L. G. (2011). Capture in Financial Regulation: Can We Channel It toward the Common Good. *Cornell Journal of Law and Public Policy*, *211*(January), 175–200.

Bevan, G., & Hood, C. (2006). What's measured is what matters: Targets and gaming in the English public health care system. *Public Administration*, *84*(3), 517–538. https://doi.org/10.1111/j.1467-9299.2006.00600.x

Bijker, W. E., & Law, J. (1992). *Shaping technology/building society : studies in sociotechnical change*. Massachusetts: MIT Press.

Bruno, I., Jacquot, S., & Mandin, L. (2006). Europeanization through its instrumentation: benchmarking, mainstreaming and the open method of co-ordination … toolbox or Pandora's box? *Journal of European Public Policy*, *13*(4), 519–536. https://doi.org/10.1080/13501760600693895

Bunea, A. (2013). Issues, preferences and ties: determinants of interest groups' preference attainment in the EU environmental policy. *Journal of European Public Policy*, *20*(4), 552–570. https://doi.org/10.1080/13501763.2012.726467

Callon, M. (1984). Some Elements of a Sociology of Translation: Domestication of the Scallops and the Fishermen of St Brieuc Bay. *The Sociological Review*, *32*(1_suppl), 196–233. https://doi.org/10.1111/j.1467-954X.1984.tb00113.x

Cecchetti, S. G., & Schoenholtz, K. L. (2016). Are European Stress Tests Stressful Enough?

Moneyandbanking.Com.

Dal Bo, E. (2006). Regulatory Capture: A Review. *Oxford Review of Economic Policy*, *22*(2), 203–225. https://doi.org/10.1093/oxrep/grj013

Davis, K. E., Kingsbury, B., & Merry, S. E. (2012). Indicators as a Technology of Global Governance. *Law & Society Review*, *46*(1), 71–104.

De Bruyckere, V., Gerhardt, M., Schepens, G., & Vander Vennet, R. (2013). Bank/sovereign risk spillovers in the European debt crisis. *Journal of Banking & Finance*, *37*(12), 4793–4809. https://doi.org/10.1016/J.JBANKFIN.2013.08.012

Desrosières, A. (1998). *The politics of large numbers : a history of statistical reasoning*. Harvard University Press.

EBA. (2014). Results - European Banking Authority. Retrieved February 27, 2018, from https://www.eba.europa.eu/risk-analysis-and-data/eu-wide-stress-testing/2014/results

Espeland, W. N., & Sauder, M. (2007). Rankings and Reactivity: How Public Measures Recreate Social Worlds. *American Journal of Sociology*, *113*(1), 1–40. https://doi.org/10.1086/517897

Espeland, W. N., & Stevens, M. L. (1998). Commensuration as a social process. *Annual Review of Sociology*, *24*, 313–43.

Etzioni, A. (2009). The Capture Theory of Regulations—Revisited. *Society*, *46*(4), 319–323. https://doi.org/10.1007/s12115-009-9228-3

Gorur, R. (2016a). Seeing like PISA: A cautionary tale about the performativity of international assessments. *European Educational Research Journal*, *15*(5), 598–616. https://doi.org/10.1177/1474904116658299

Gorur, R. (2016b). Seeing like PISA: A cautionary tale about the performativity of international assessments. *European Educational Research Journal*. https://doi.org/10.1177/1474904116658299

Grabosky, P., & Braithwaite, J. (1986). *Of Manners Gentle - Enforcement strategies of Australian business regulatory agencies*. Melbourne: Oxford University Press.

Habermas, J. (1984). *The theory of communicative action, volume one: Reason and the rationalization of society*. Boston: Beacon Press . Retrieved from http://www.dphu.org/uploads/attachements/books/books_2795_0.pdf

Habermas, J. (1990). *Moral consciousness and communicative action*. MIT Press.

Habermas, J., & Derrida, J. (2003). February 15, or What Binds Europeans Together: A Plea for a Common Foreign Policy, Beginning in the Core of Europe. *Constellations*, *10*(3), 291–297. https://doi.org/10.1111/1467-8675.00333

Hacking, I. (1990). *The taming of chance*. Cambridge Univeristy Press.

Hanegraaff, M., Beyers, J., & De Bruycker, I. (2016). Balancing inside and outside lobbying: The political strategies of lobbyists at global diplomatic conferences. *European Journal of Political Research*, *55*(3), 568–588. https://doi.org/10.1111/1475-6765.12145

Hood, C., & Peters, G. (2004). The Middle Aging of New Public Management: Into the Age of Paradox? *Journal of Public Administration Research and Theory*, *14*(3), 267–282. https://doi.org/10.1093/jopart/muh019

Horkheimer, M., & Adorno, T. W. (1947). *Dialektik der Aufklärung: Philosophische Fragmente*. Amsterdam: Querido.

Huault, I., & Rainelli-Weiss, H. (2011). A Market for Weather Risk? Conflicting Metrics, Attempts at Compromise, and Limits to Commensuration. *Organization Studies*, *32*(10), 1395–1419. https://doi.org/10.1177/0170840611421251

Jasanoff, S. (1986). *Risk Management and Political Culture: A Comparative Study of Science in the Policy Context*. (S. Jasanoff, Ed.) (1st ed.). New York: Russel Sage Foundation.

Klüver, H. (2013). Lobbying as a collective enterprise: winners and losers of policy formulation in the European Union. *Journal of European Public Policy*, *20*(1), 59–76. https://doi.org/10.1080/13501763.2012.699661

Kohler-Koch, B. (1996). Catching up with change: The transformation of governance in the European Union. *Journal of European Public Policy*, *3*(3), 359–380. https://doi.org/10.1080/13501769608407039

Kolk, A., Levy, D., & Pinkse, J. (2008). Corporate Responses in an Emerging Climate Regime: The Institutionalization and Commensuration of Carbon Disclosure. *European Accounting Review*, *17*(4), 719–745. https://doi.org/10.1080/09638180802489121

Latour, B. (1987). *Science in action: How to follow scientists and engineers through society*. Harvard University Press.

Latour, B. (1999). On recalling ANT. *The Sociological Review*, *47*(1), 15–25. https://doi.org/10.1111/j.1467-954X.1999.tb03480.x

Latour, B., & Woolgar, S. (n.d.). *Laboratory Life: The Construction of Scientific Facts*. Princeton University Press.

Law, J. (1999). After Ant: Complexity, Naming and Topology. *The Sociological Review*, *47*(1_suppl), 1–14. https://doi.org/10.1111/j.1467-954X.1999.tb03479.x

Lowery, D. (2013). Lobbying influence: Meaning, measurement and missing. *Interest Groups & Advocacy*, *2*(1), 1–26. https://doi.org/10.1057/iga.2012.20

Marlier, E., & Atkinson, A. B. (2010). Indicators of poverty and social exclusion in a global context. *Journal of Policy Analysis and Management*, *29*(2), 285–304. https://doi.org/10.1002/pam.20492

McCarthy, T. (1990). The Critique of Impure Reason. *Political Theory*, *18*(3), 437–469. https://doi.org/10.1177/0090591790018003005

Mügge, D. (2016). Studying macroeconomic indicators as powerful ideas. *Journal of European Public Policy*, *23*(3), 410–427. https://doi.org/10.1080/13501763.2015.1115537

Patterson, M. (1998). Commensuration and theories of value in ecological economics. *Ecological Economics*, *25*(1), 105–125. https://doi.org/10.1016/S0921-8009(97)00166-3

Peeters, H., & Verschraegen, G. (3013). Governance by numbers: risico's verbonden aan de internationale benchmarking en ranking van pensioensystemen *. *Beleid En Maatschappij*, *40*(2), 133–155.

Peeters, H., Verschraegen, G., & Debels, A. (2014). Commensuration and policy comparison: How the use of standardized indicators affects the rankings of pension systems. *Journal of European Social Policy*, *24*(1), 19–38. https://doi.org/10.1177/0958928713511279

Politt, C., & Talbot, C. (2004). *Unbundled Government: A Critical Analysis of the Global Trend to Agencies*. London: Routledge.

Porter, T. M. (1995). *Trust in numbers : the pursuit of objectivity in science and public life*. Princeton University Press.

Power, M. (2003). Auditing and the production of legitimacy. *Accounting, Organizations and Society*, *28*(4), 379–394. https://doi.org/10.1016/S0361-3682(01)00047-2

Rose, N. (1991). Governing by numbers: Figuring out democracy. *Accounting, Organizations and Society*, *16*(7), 673–692. https://doi.org/10.1016/0361-3682(91)90019-B

Rottenburg, R., & Merry, S. E. (2015). A world of indicators: The making of governmental knowledge through quantification. In R. Rottenburg, S. E. Merry, S.-J. Park, & J. Mugler (Eds.), *The World of Indicators* (pp. 1–33). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781316091265.001

Scott, J. C. (1998). *Seeing like a state : how certain schemes to improve the human condition have failed*. Yale University Press.

Smith, P. (1995). Performance indicators and outcome in the public sector. *Public Money & Management*, *15*(4), 13–16. https://doi.org/10.1080/09540969509387889

Smulewicz-Zucker, G. (2017). The Frankfurt School and the Critique of Instrumental Reason. In M. Thopson (Ed.), *The Palgrave Handbook of Critical Theory. Political Philosophy and Public Purpose.* (pp. 185–206). New York: Palgrave Macmillan.

Tervonen-Gonçalves, L. (2012). From averages to best performers: use of comparisons in

identity formation. *Critical Policy Studies*, *6*(3), 304–323. https://doi.org/10.1080/19460171.2012.717784

Titcomb, J. (2014). Eurozone banks rush to fill black hole ahead of stress tests - Telegraph. Retrieved February 27, 2018, from https://www.telegraph.co.uk/finance/newsbysector/banksandfinance/11155339/Eurozone-banks-rush-to-fill-black-hole-ahead-of-stress-tests.html

van Ostaijen, M., & Scholten, P. (2017). The politics of numbers. Framing intra-EU migrants in the Netherlands. *Critical Policy Studies*, *11*(4), 477–498. https://doi.org/10.1080/19460171.2016.1224725

Vickerman, R. (2007). Cost — Benefit Analysis and Large-Scale Infrastructure Projects: State of the Art and Challenges. *Environment and Planning B: Planning and Design*, *34*(4), 598–610. https://doi.org/10.1068/b32112

Welsh, J. (2017). Ranking academics: toward a critical politics of academic rankings. *Critical Policy Studies*, 1–21. https://doi.org/10.1080/19460171.2017.1398673

Woll, C. (2009). Who Captures Whom? Trade Policy Lobbying in the European Union, 268–288. Retrieved from https://hal-sciencespo.archives-ouvertes.fr/hal-00972851/

Woll, C. (2014). *The power of inaction : bank bailouts in comparison*.