Review of International Studies

Special Issue: The Politics of Numbers
Guest Editors: André Broome and Joel Quirk

Editors

Kimberly Hutchings
Mathias Albert
George Lawson
Jennifer Sterling-Folker

Review of International Studies

© British International Studies Association 2015

Review of International Studies

CONTENTS

The politics of numbers: the normative agendas of global benchmarking

ANDRÉ BROOME AND JOEL QUIRK*

Abstract. Global benchmarks have grown exponentially over the last two decades, having been both applied to and developed by states, international organisations, corporations, and non-governmental organisations. As a consequence, global benchmarking is now firmly established as a distinct mode of transnational governance. Benchmarking chiefly involves the development of comparative metrics of performance, which typically take the form of highly stylised comparisons which are generated by translating complex phenomena into numerical values via simplification and extrapolation, commensuration, reification, and symbolic judgements. This process of translation takes what might otherwise be highly contentious normative agendas and converts them into formats that gain credibility through rhetorical claims to neutral and technocratic assessment. This politics of numbers has far-reaching ramifications for transnational governance, including the dimensions and effects of indirect power, expertise and agenda-setting, coordination, regulation and certification, and norm contestation and activism. This Special Issue draws upon an emerging literature to explore how and why benchmarks both align with and expand upon established models of International Relations theory and scholarship. It does so by critically examining the role of global benchmarks in key areas such as state 'failure', global supply chains, disaster management, economic governance, corporate social responsibility, and human development.

André Broome is Associate Professor of International Political Economy in the Department of Politics and International Studies and is Director of the Centre for the Study of Globalisation and Regionalisation, University of Warwick. His publications include *The Currency of Power* (Palgrave, 2010) and *Issues and Actors in the Global Political Economy* (Palgrave, 2014).

Joel Quirk is an Associate Professor in the Department of Political Studies, University of the Witwatersrand. His publications include *The Anti-Slavery Project* (Penn, 2011) and *Mobility Makes States* (Penn, 2015). Joel is a current member of the International Scientific Committee of the UNESCO Slave Route Project, and is also an *openDemocracy* editor for Beyond Trafficking and Slavery.

The political appeal of numbers

In November 2014, the Walk Free Foundation launched the second edition of their flagship Global Slavery Index. The press release for the launch declared the index to be 'the most accurate and comprehensive measure of the extent and risk of modern slavery', building upon an 'improved methodology' which included random-sample

surveys from 19 countries.[1] This was significant because the inaugural 2013 index had been heavily criticised for using unreliable, incomplete, and inappropriate data.[2] Despite this emphasis on improved methods, the second version of the index once again attracted sustained criticism. One of the sharpest challenges came from Anne Gallagher, a noted expert on human trafficking, who reported that the index contained 'critical errors of fact and logic'.[3]

The errors which Gallagher and others have identified chiefly relate to the steps which are required in order to generate a large volume of numerical information describing practices through the globe. Complex social, economic and political phenomenon, such as state stability and discrimination, are made both easily accessible and globally commensurable through a process of radical simplification and 'guesstimation'. Distinctive qualities are converted into numerical quantities, which are then compared and assessed in terms of orders of magnitude. Challenging and contested concepts, such as slavery, acquire fixed and unproblematic meanings, which are presumed to be universally applicable irrespective of cultural context. Once it has been crudely converted into numerical values, the Global Slavery Index then goes on to assign countries a ranking, with a ranking of 1 being the worst and a ranking of 167 being the best. This also extends to a 'Top Ten' list, which is a popular device used to draw attention to the 'best' or 'worst' performers. In a world where school and university league tables, university rankings and other metrics have become pervasive, the idea of ranking countries based on performance is a concept we are already very familiar with.

The Global Slavery Index is not a new innovation, but can instead be best understood as the extension of an already well-established template to a new topic. Over the last two decades, there has been a remarkable explosion in the prevalence of national, regional, and global benchmarks. Some notable examples from a much larger trend include the Fragile/Failed States Index (Fund for Peace, from 2005), the Corruption Perceptions Index (Transparency International, 1995), the Climate Change Performance Index (Germanwatch & Climate Action Network Europe, 2006), and the more venerable Freedom in the World report (Freedom House, 1972), the Trafficking in Persons Report (Department of State, United States, 2001), the Human Development Index (United Nations Development Programme, 1990), the Foreign Direct Investment Confidence Index (A. T. Kearney, 1998), and the Bloomberg Innovation Index (Bloomberg, 2013). It is hard to think of any area of International Relations, from international security to global political economy, grand strategy, climate change, human rights, international development, and global public policy, which has not been pulled into this politics of numbers. As this Special Issue demonstrates, this recent popularity can be chiefly traced to the capacity of global benchmarks to cloak normative agendas in languages of neutral and technocratic assessment.

---

[1] '35.8 Million People are Enslaved Across the World', *Global Slavery Index* (5 November 2014), available at: {http://www.globalslaveryindex.org/35-8-million-people-are-enslaved-across-the-world/} accessed 29 July 2015.

[2] Andrew Guth, Robyn Anderson, Kasey Kinnard, and Hang Tran, 'Proper methodology and methods of collecting and analyzing slavery data: an examination of the Global Slavery Index', *Social Inclusion*, 2:4 (2014), pp. 14–22; Ronald Weitzer, 'Miscounting human trafficking and slavery', *openDemocracy*, available at: {https://www.opendemocracy.net/beyondslavery/ronald-weitzer/miscounting-human-trafficking-and-slavery}; Neil Howard, 'Keeping Count: The Trouble with the Global Slavery Index', *The Guardian* (13 January 2014), available at: {http://www.theguardian.com/global-development-professionals-network/2014/jan/13/slavery-global-index-reports} accessed 29 July 2015.

[3] Anne Gallagher, 'The Global Slavery Index is Based on Flawed Data – Why Does No One Say So?', *The Guardian* (28 November 2014), available at: {http://www.theguardian.com/global-development/poverty-matters/2014/nov/28/global-slavery-index-walk-free-human-trafficking-anne-gallagher} accessed 29 July 2015.

The normative agendas of global benchmarking

Global benchmarking involves the classification of relative performance or value. In this Special Issue we use *global benchmarking* as an umbrella term for a wide range of comparative evaluation techniques that systematically assess the performance of actors, populations, or institutions. This can include techniques such as audits, rankings, indicators, indexes, baselines, and targets, all of which work on the basis of standardised measurements, metrics, and rankings. We define global benchmarking as a specifically 'transnational' practice that goes beyond the jurisdictions of individual states, although particular benchmarks may not be fully 'universal' in scope. Benchmarking is further understood as encompassing one or more of the following forms of comparative assessment: (1) quality of *conduct*, or how well actors have discharged their responsibilities in specific areas; (2) quality of *design*, or how well specific policies, laws, or institutions have been formulated and applied; and (3) quality of *outcomes*, or how well activities in specific areas align with defined goals (irrespective of who is actually responsible for the overall outcomes).

The main focus in the Special Issue is on benchmarking by external transnational actors, rather than internal self-benchmarking by states or corporations.[4] As we discuss further in the lead article for the Issue,[5] the recent proliferation of global benchmarks has been driven by the political appeal of numbers as information shortcuts that are frequently assumed to present unbiased facts.[6] This is a recursive process whereby complex and contested normative values are translated into simplified numerical representations, which in turn enables global benchmarks to be represented as 'evidence' that can be used to establish a foundation for initiating particular kinds of political conservations as well as potentially influencing the design of policy interventions and reforms. This process of numerical translation helps to obscure the normative values that underlie global benchmarks, and also enables non-experts to make simplistic comparisons of relative performance regarding complex phenomena at a transnational level.

We suggest that this process of numerical translation is common to all forms of global benchmarking, and can be divided into four (non-sequential) components: (1) simplification and extrapolation; (2) commensuration; (3) reification; and (4) symbolic judgement.[7] There is a distinct form of the 'politics of numbers' at work here. In the short term, translating qualities into quantities can potentially expand the political traction that global benchmarks achieve, by increasing their salience at the same time as it simplifies the message they communicate. In the longer

[4] See, for example, Sakiko Fukuda-Parr, 'Millennium development Goal 8: Indicators for international human rights obligations?', *Human Rights Quarterly*, 28:4 (2006), pp. 966–7; Tore Fougner, 'Neoliberal governance of states: the role of competitiveness indexing and country benchmarking', *Millennium: Journal of International Studies*, 37:2 (2008), pp. 303–26; Peter Larmour, 'Civilizing techniques: Transparency international and the spread of anti-corruption', in Brett Bowden and Leonard Seabrooke (eds), *Global Standards of Market Civilization* (Abingdon: Routledge, 2006), pp. 95–106; Laura Langbein and Stephen Knack, 'The worldwide governance indicators: Six, one, or none?', *Journal of Development Studies*, 46:2 (2010), pp. 350–70; Nehal Bhuta, 'Governmentalizing sovereignty: Indexes of state fragility and the calculability of political order', in Kevin E. Davis, Angelina Fisher, Benedict Kingsbury, and Sally Engle Merry (eds), *Governance by Indicators: Global Power through Quantification* (Oxford: Oxford University Press, 2012), pp. 132–62.
[5] See André Broome and Joel Quirk, 'Governing the world at a distance: the practice of global benchmarking', *Review of International Studies*, 41:5 (2015), pp. 819–41.
[6] Lorenzo Fioramonti, *How Numbers Rule the World: The Use and Abuse of Statistics in Global Politics* (London: Zed Books, 2014), p. 192.
[7] For a full explanation of these concepts, see Broome and Quirk, 'Governing the world at a distance'.

term, global benchmarks can potentially create 'anchoring effects' by establishing referents that shape how different actors subsequently think about and 'see' specific issues.[8] Different forms of global benchmarking may sometimes operate in competition, especially when benchmark producers are seeking to construct issue expertise or to otherwise compete for authority in a given area.[9] Because individual global benchmarks often share a common normative orientation, however, they therefore contribute to the diffusion of overlapping normative agendas regarding what particular kinds of transnational actors should look like, what they should value, and how they should behave.

Towards a new IR research agenda on global benchmarking

Global benchmarking represents a new and distinctive application of authority in world politics. Benchmarking differs from earlier efforts to promulgate a 'standard of civilization' based on Western values,[10] as well as more recent attempts to diffuse global 'best practice' policy norms through international regimes and organisations.[11] Because the quantification of particular norms and values relies primarily on the imprimatur of scientific expertise for establishing authority, and because this authority to make symbolic judgements is typically expressed through the language of numbers, global benchmarks are both easier to promulgate and, potentially, more difficult for target actors to ignore. It may be harder to argue with a numerical ranking, especially when this serves as a reference point for political debates about a given issue, than it is to resist technical advice from international actors on complex institutional reforms, except through deploying rival expertise and challenging the credibility of a global benchmark on the same ontological terrain.

Collectively, the nine articles in this Special Issue represent an important step in advancing the theorisation and empirical analysis of the politics of numbers and the dynamics of global benchmarking in IR. In the lead article for the Issue we seek to place global benchmarking on the IR agenda in accessible and general terms through developing a conceptual apparatus for the study of benchmarking, and by mapping out a fourfold typology for distinguishing between different types of benchmarking as a transnational practice.[12] This is followed by Alexandra Homolar's article on human security benchmarks, which provides a robust critical examination of how the role of benchmarking within the wider human security agenda has served to reinforce the state as the main focal point of international security governance. Homolar's analysis turns much of the conventional wisdom about human security in IR on its head, and shows how the emancipatory potential often associated with the shift towards the human security agenda has been undermined. As a consequence, global humanitarian

---

[8] Peter Andreas and Kelly M. Greenhill, 'Introduction: the politics of numbers', in Peter Andreas and Kelly M. Greenhill (eds), *Sex, Drugs, and Body Counts: The Politics of Numbers in Global Crime and Conflict* (Ithaca, NY: Cornell University Press, 2010), p. 17; see also André Broome and Leonard Seabrooke, 'Seeing like an international organisation', *New Political Economy*, 17:1 (2012), pp. 1–16.

[9] See Leonard Seabrooke, 'Epistemic arbitrage: Transnational professional knowledge in action', *Journal of Professions and Organizations*, 1:1 (2014), pp. 49–64; Ole Jacob Sending, *The Politics of Expertise: Competing for Authority in Global Governance* (Ann Arbor: University of Michigan Press, 2015).

[10] See Shogo Suzuki, *Civilization and Empire: China and Japan's Encounter with European International Society* (Abingdon: Routledge, 2009).

[11] See Broome and Seabrooke, 'Seeing like an international organisation'.

[12] Using this typology, we have compiled a Global Benchmarking Database consisting of 205 benchmarks (as of June 2015), which is available at: {www.warwick.ac.uk/globalbenchmarking/database}.

governance has been inscribed with new power relations centred on the attempt to fix a particular conceptualisation of a 'secure' human life.

Tony Porter's examination of global benchmarking networks helps to unpack the organisational contexts of benchmarking. Using actor-network theory, Porter provides a systematic analysis of two key and underexplored features of global benchmarking: network embeddedness, and network publicness. The article demonstrates how the logic of benchmarking can create an impetus towards increased publicness, whereby benchmarking networks are characterised by expanding flows of information across widening circles of engagement. Rather than simply focussing on the power of benchmark producers, the relative weakness of target actors, or the formal design of benchmarking metrics, Porter makes an important and compelling argument of the need for shifting the focus of analysis towards the organisational contexts in which benchmarking networks operate.

In their investigation of how activists use benchmarking initiatives, Leonard Seabrooke and Duncan Wigan characterise benchmarks as a form of 'symbolic violence' that NGOs utilise as weapons to exert pressure on a range of transnational actors, such as firms, states, and international organisations. Drawing on the work of Pierre Bourdieu, Seabrook and Wigan's analysis of the role of benchmarking in transnational advocacy strategies helps to explain three constituent elements for successful benchmarking by NGOs: salience, will, and expertise. Differentiating between two types of benchmarking cycles, 'reformist benchmarking' and 'revolutionary benchmarking', Seabrooke and Wigan's contribution to the Issue helps to shed new light on the conditions under which activist organisations' campaigns on complex issues can be mounted and sustained using benchmarking practices.

Genevieve LeBaron and Jane Lister's article critically evaluates the power relations inherent in the 'ethical' compliance audit regime, focusing on the effectiveness of supply chain benchmarking as a mechanism for promoting environmental and social improvements in global retail supply chains. As LeBaron and Lister's research reveals, while benchmarks and audits have become increasingly trusted forms of corporate governance by governments and the public, they can function to obscure serious problems in global supply chains due to selective coverage and the challenges inherent in regulating labour rights and safeguards through numbers compared with capital and product quality. Within the context of a broader trend towards increasing reliance on global supply chain benchmarking as a form of transational corporate governance, LeBaron and Lister's article makes an important intervention that highlights the potential for benchmarking to be used to expand corporate power over consumers as well as national policymakers, and helps to focus attention on the dangers inherent in corporate 'self-regulation' at the transnational level.

James Harrison and Sharifah Sekalala's article critically examines how global benchmarking initiatives by the United Nations (UN) help to promote compliance with international human rights norms. Through a comparative analysis of benchmarks produced by the UN for states and those produced for corporations, Harrison and Sekalala reveal the limits of global benchmarking as an effective mode of transnational governance. In particular, they explore how global benchmarking initiatives can be used to provide 'superficial legitimation' for the human rights performance of corporations. Rather than helping to close 'the compliance gap' identified in the implementation of international human rights norms, Harrison and Sekalala show how benchmarks based on self-reporting initiatives can instead create opportunities for powerful actors to evade their obligations by gaming the system.

Liam Clegg investigates the political contestation of how the success of the Millennium Development Goal (MDG) framework was benchmarked in UN reviews over the last fifteen years. Clegg shows that reviews by the UN of the progress achieved under the MDG framework was riven by a series of 'blame games' over how responsibility for poor performance was attributed. He identifies a clear split between developed and developing countries, with the former blaming internal governance failings in developing countries for poor performance, while the latter blamed aid shortfalls for a lack of success in achieving MDG targets. As Clegg's analysis shows, the problems caused by the imbrication of benchmarking and blame games over how responsibility is attributed for the success or failure of global development initiatives are likely to be carried over into the design of the new Sustainable Development Goals.

Caroline Kuzemko's analysis of European Union (EU) climate change benchmarking practices reveals the competing motivations that lie behind the development of climate governance, and shows how climate change benchmarking practices function to increase the legibility of certain political and economic practices, while obscuring others. Through a comparative analysis of two cases of 'best practice' countries – the United Kingdom and Germany – Kuzemko demonstrates how the quantification processes and short time horizons inherent in climate change benchmarking can give a distorted picture of domestic compliance with EU climate change initiatives. As a consequence, Kuzemko argues that climate change benchmarking often serves to gloss over important political issues for long-term climate mitigation, while simultaneously side-lining more radical policy alternatives.

The final article in the Issue, by Ole Jacob Sending and Jon Harald Sande Lie, critically examines how the World Bank benchmarks African economies through its Country Policy and Institutional Assessment ratings. Sending and Lie show how benchmarking connects policy and practice. In particular, their analysis shows that while benchmarks may be used to police and shape target actors' behaviour in some cases, in others they are used to manage and produce existing relations of power, thereby constituting particular forms of authority between actors based on their respective identities.

The key conceptual themes developed throughout the Special Issue therefore centre on the role of global benchmarking practices in attributing responsibility for good or bad performance, constructing issue expertise by both state and non-state actors, altering the terms of political rhetoric, shaping formal policy agendas, and defining the concepts used to quantify international norms. Many of the contributions also provide new insights into how global benchmarking practices may constitute new sets of relations between different types of transnational actors, and explore the role of 'third parties' who use benchmarks produced by others for a variety of political purposes. Covering a wide range of rich empirical terrain, the contributions to this Issue help to expand our understanding of how global benchmarking practices have emerged as a critical mode of transnational governance in the twenty-first century.

Governing the world at a distance: the practice of global benchmarking

ANDRÉ BROOME AND JOEL QUIRK*

Abstract. Benchmarking practices have rapidly diffused throughout the globe in recent years. This can be traced to their popularity amongst non-state actors, such as civil society organisations and corporate actors, as well as states and international organisations (IOs). Benchmarks serve to both 'neutralise' and 'universalise' a range of overlapping normative values and agendas, including freedom of speech, democracy, human development, environmental protection, poverty alleviation, 'modern' statehood, and 'free' markets. The proliferation of global benchmarks in these key areas amounts to a comprehensive normative vision regarding what various types of transnational actors should look like, what they should value, and how they should behave. While individual benchmarks routinely differ in terms of scope and application, they all share a common foundation, with normative values and agendas being translated into numerical representations through simplification and extrapolation, commensuration, reification, and symbolic judgements. We argue that the power of benchmarks chiefly stems from their capacity to create the appearance of authoritative expertise on the basis of forms of quantification and numerical representation. This politics of numbers paves the way for the exercise of various forms of indirect power, or 'governance at a distance', for the purposes of either status quo legitimation or political reform.

André Broome is Associate Professor of International Political Economy in the Department of Politics and International Studies and Director of the Centre for the Study of Globalisation and Regionalisation, University of Warwick. His publications include *The Currency of Power* (Palgrave, 2010) and *Issues and Actors in the Global Political Economy* (Palgrave, 2014).

Joel Quirk is Associate Professor in the Department of Political Studies, University of the Witwatersrand. His publications include *The Anti-Slavery Project* (Penn, 2011) and *Mobility Makes States* (Penn, 2015). Joel is a current member of the International Scientific Committee of the UNESCO Slave Route Project, and is also an *openDemocracy* editor for Beyond Trafficking and Slavery.

Introduction

Global benchmarking comprises a distinct type of transnational practice in contemporary world politics, which involves the development and application of

comparative metrics of performance. While benchmarking is not in itself a new phenomenon, the last three decades have been marked by a sharp increase in the density, complexity, and coverage of global benchmarking practices.[1] Much of this ongoing trend can be traced to the globalisation of an 'audit explosion' that began in the 1980s in domestic political contexts, and which has had far-reaching ramifications for both public and private processes of transnational governance.[2] Other key contributing factors include the rapid proliferation of non-governmental organisations (NGOs) in areas such as human rights, health, gender, and the environment, together with a parallel shift from state to private regulation at a corporate level.[3] Even the 'ivory tower' of academia is increasingly governed through ratings, rankings, and measurements of how well higher education institutions perform in comparison to their competitors.[4] These and other developments have not only dramatically expanded the pool of prospective 'benchmarkers'. They have also fostered an environment where benchmarks have gained considerable legitimacy and authority.

In its most basic form, benchmarking involves the classification of relative performance or value. In this article and for the Special Issue, benchmarking is used as an umbrella term for a wide range of comparative evaluation techniques – such as audits, rankings, indicators, indexes, baselines, or targets – which systematically assess the performance of actors, populations, or institutions on the basis of standardised measurements, metrics, and rankings. More specifically, benchmarking involves one or more of the following forms of comparative assessment: (1) quality of *conduct*, or how well actors have discharged their responsibilities in specific areas; (2) quality of *design*, or how well specific policies, laws, or institutions have been formulated and applied; and (3) quality of *outcomes*, or how well activities in specific areas align with defined goals (irrespective of who is actually responsible for the overall outcomes).

In this article we identify and analyse a number of core features of benchmarking as a distinct mode of governance in world politics. We begin our analysis by locating global benchmarking within an emerging literature that focuses on how and why both states and non-state actors have sought to regulate and shape transnational issues through indirect forms of power, rather than through direct compulsion. Building upon this literature, we argue that benchmarking can be best understood as an exercise in 'governing at a distance', wherein the power of benchmarks primarily stems from their capacity to indirectly shape procedural standards, issue expertise, institutional

---

[1] See the Global Benchmarking Database (N = 205), version 1.8, available at: {www.warwick.ac.uk/globalbenchmarking/database} accessed 5 June 2015. We are grateful to Matthias Kranke for research assistance with compiling data on Global Benchmarks.

[2] See, for example, Michael Power, *The Audit Society: Rituals of Verification* (Oxford: Oxford University Press, 1997); Wendy Nelson Espeland and Mitchell L. Stevens, 'A sociology of quantification', *European Journal of Sociology*, 49:3 (2008), pp. 401–36; Michael Power, 'Evaluating the audit explosion', *Law and Policy*, 25:3 (2003), pp. 185–202.

[3] See, for example, Richard Locke, *The Promise and Limits of Private Power: Promoting Labor Standards in a Global Economy* (Cambridge: Cambridge University Press, 2013); Anne-Marie Slaughter, *A New World Order* (Princeton: Princeton University Press, 2005); Samuel Moyn, *The Last Utopia: Human Rights in History* (Cambridge, MA: Harvard University Press, 2012); Tim Büthe and Walter Mattli, *The New Global Rulers: The Privatization of Regulation in the World Economy* (Princeton: Princeton University Press, 2011); Angelina Fisher, 'From diagnosing under-immunization to evaluating health care systems: Immunization coverage indicators as a technology of global governance', in Kevin E. Davis, Angelina Fisher, Benedict Kingsbury, and Sally Engle Merry (eds), *Governance by Indicators: Global Power through Quantification* (Oxford: Oxford University Press, 2012), pp. 217–46.

[4] Jack Snyder and Alexander Cooley, 'Conclusion: Rating the ratings craze: From consumer choice to public policy outcomes', in Alexander Cooley and Jack Snyder (eds), *Ranking the World: Grading States as a Tool of Global Governance* (Cambridge: Cambridge University Press, 2015), pp. 180–2.

obligations, and political conversations. Much of the power of benchmarking is bound up in the mechanics and effects of ranking and quantification, which in turn generate a form of 'constructed objectivity' that acts back upon the reality it aims to describe.[5] The recent popularity of benchmarks can also be traced to their capacity to promote otherwise highly contentious policy goals and political agendas by means of rhetorical appeals to the ostensibly neutral language of technocratic assessment and numerical comparison. Complex social phenomena become legible by means of quantification, extrapolation, and simplification. Concepts such as freedom, development, and democracy, which academics routinely describe as essentially contested, instead appear as fixed, unproblematic, and reified categories.

We have divided this article into four main sections. The first section briefly situates our approach to global benchmarking within the larger context of existing literatures in International Relations (IR) on political activism and norms, rational design and institutions, and governmentality and expertise. In the second section, we focus upon the mechanics and effects associated with translating normative values into numerical representations. By radically reducing issue complexity, benchmarks have the potential to alter 'how people think about things and how information moves around the world'.[6] This process of translation can be divided into a series of steps common to all forms of benchmarking: simplification and extrapolation, commensuration, reification, and symbolic judgment. The third section examines the political ramifications of these processes of quantification and numerical representation for transnational governance, along with the political impact of the alignment between benchmarks and other agendas. The final section, which introduces a typology of global benchmarking practices, develops this line of inquiry further. We divide global benchmarking practices into four main categories: (1) statecraft; (2) international governance; (3) private market governance; and (4) transnational advocacy. We conclude the article by identifying a series of core questions for a new research agenda on global benchmarking in International Relations.

Governing at a distance: Benchmarking and IR theory

We understand global benchmarking as a mode of *transnational governance*, which comprises a patchwork of political structures within and above the state that envelop, constrain, and enable various actors. Drawing on Marie-Laure Djelic and Kerstin Sahlin-Andersson's definition, the boundaries of the 'transnational' arena stretch beyond the jurisdiction of domestic governance structures and are not limited to one specific region.[7] Benchmarking practices are *global* when they aim to produce comparative measurements of performance across numerous countries and regions. The units of analysis comprise a range of transnational actors, such as states,

---

[5] Peter Berger and Thomas Luckmann, *The Social Construction of Reality: A Treatise in the Sociology of Knowledge* (London: Allen Lane, 1967), p. 78.

[6] Wendy Nelson Espeland and Michael Sauder, 'The dynamism of indicators', in Kevin E. Davis, Angelina Fisher, Benedict Kingsbury, and Sally Engle Merry (eds), *Governance by Indicators: Global Power through Quantification* (Oxford: Oxford University Press, 2012), p. 91.

[7] Marie-Laure Djelic and Kerstin Sahlin-Andersson, 'Introduction: a world of governance – the rise of transnational regulation', in Marie-Laure Djelic and Kerstin Sahlin-Andersson (eds), *Transnational Governance: Institutional Dynamics of Regulation* (Cambridge: Cambridge University Press, 2008), p. 4; see also Emanuel Adler and Vincent Pouliot, 'International practices: Introduction and framework', in Emanuel Adler and Vincent Pouliot (eds), *International Practices* (Cambridge: Cambridge University Press, 2011), pp. 7–8.

international organisations, or corporate subsidiaries within a global production network. A global framework applies even if some countries or actors are excluded.

Many global benchmarking efforts have focused on the economic and political performance of states. Early examples on the economic front include gross domestic product and the System of National Accounts, developed during the 1930s in the United States.[8] With respect to political performance, pioneering examples of benchmarking include the standardised international monitoring of elections and the annual 'Freedom in the World' rankings published by Freedom House (an NGO part-funded by the US government) since 1973.[9] In addition to measuring cross-national economic and political performance, benchmarking has also become an important means for evaluating corporate performance. This involves systematic comparisons to evaluate individual firm competitiveness and to establish industry 'best practice' processes based on measures of quality, time, and cost.[10] This form of benchmarking extends to commercially motivated efforts to evaluate market conditions, financial performance, and creditworthiness, most notably by means of credit ratings. In some cases political and commercial concerns have been integrated, such as in the political and country risk ratings published by the PRS Group since 1980.[11] One of the distinctive features of corporate benchmarking is that it frequently takes the form of self-benchmarking against peers with a view to improving, validating, or refining overall performance and internal processes,[12] which is broadly comparable to the use of benchmarking by individual states for the purposes of domestic governance.

Our main focus here is on benchmarking by external transnational actors, rather than internal self-benchmarking. Some notable examples of this trend include measures of state performance in relation to international human rights obligations,[13] global indexes of country 'competitiveness',[14] measurements of the perception of corruption in state institutions,[15] assessments of democratic freedom and the

---

[8] Yoshiko M. Herrera, *Mirrors of the Economy: National Accounts and International Norms in Russia and Beyond* (Ithaca, NY: Cornell University Press, 2010); Lorenzo Fioramonti, *Gross Domestic Problem: The Politics Behind the World's Most Powerful Number* (London: Zed Books, 2013); Daniel Mügge, 'Fickle formulas: Towards a political economy of macroeconomic measurements', *Journal of European Public Policy*, forthcoming.

[9] Judith G. Kelley, *Monitoring Democracy: When International Election Observation Works, and Why It Often Fails* (Princeton: Princeton University Press, 2012); see Alexandra Homolar, 'Human security benchmarks: Governing human wellbeing at a distance', *Review of International Studies*, 41:5 (2015), pp. 843–63.

[10] Wendy Larner and Richard Le Heron, 'Global benchmarking: Participating "at a distance" in the globalizing economy', in Wendy Larner and William Walters (eds), *Global Governmentality: Governing International Spaces* (Abingdon: Routledge, 2004), pp. 212–32.

[11] Nikhil K. Dutta, 'Accountability in the generation of governance indicators', in Kevin E. Davis, Angelina Fisher, Benedict Kingsbury, and Sally Engle Merry (eds), *Governance by Indicators: Global Power through Quantification* (Oxford: Oxford University Press, 2012), pp. 437–64.

[12] See Genevieve LeBaron and Jane Lister, 'Benchmarking global supply chains: the power of the "ethical audit" regime', *Review of International Studies*, 41:5 (2015), pp. 905–24.

[13] Kate Raworth, 'Measuring human rights', *Ethics and International Affairs*, 15:1 (2001), pp. 111–31; Sakiko Fukuda-Parr, 'Millennium development goal 8: Indicators for international human rights obligations?', *Human Rights Quarterly*, 28:4 (2006), pp. 966–97.

[14] Tore Fougner, 'Neoliberal governance of states: the role of competitiveness indexing and country benchmarking', *Millennium: Journal of International Studies*, 37:2 (2008), pp. 303–26.

[15] Peter Larmour, 'Civilizing techniques: Transparency international and the spread of anti-corruption', in Brett Bowden and Leonard Seabrooke (eds), *Global Standards of Market Civilization* (Abingdon: Routledge, 2006), pp. 95–106; Laura Langbein and Stephen Knack, 'The worldwide governance indicators: Six, one, or none?', *Journal of Development Studies*, 46:2 (2010), pp. 350–70; Paul M. Heywood and Jonathan Rose, '"Close but no cigar": the measurements of corruption', *Journal of Public Policy*, 34:3 (2014), pp. 507–29.

transparency of elections,[16] headcount measures of absolute poverty,[17] and measures of state 'fragility'.[18] Such external benchmarking by transnational actors has rapidly proliferated around the world over the last three decades.

IR theorists have developed a number of insights and arguments that can be usefully applied in order to better understand the politics of benchmarking. Since relatively few IR theorists have focused upon benchmarking as a specific object of analysis,[19] we briefly engage with a number of allied literatures that speak to similar and related topics, most notably in relation to theories of norms and human rights, rational design and cooperation, and governmentality. Over the last two decades, IR theorists have repeatedly demonstrated that normative arguments and collective identities have generated outcomes that cannot be explained in terms of power and interest alone.[20] This has in turn resulted in sustained interest in the techniques, alliances, and arguments employed by 'agents of change'. Many of the political levers that theorists have identified – such as reputational challenge, communicative networks, and patterns of socialisation – can also be applied to the politics of benchmarking, particularly in relation to transnational advocacy. Especially relevant is the emerging literature on 'merchants of morality', which seeks to explain why and how some issues have become subject to mobilisation while others remain dormant;[21] why some political causes and organisations have secured greater success (or 'salience') than their competitors;[22] and how the accumulation and application of 'credibility' has emerged as a key source of authority and influence for NGOs.[23] Within the context of this recent literature, benchmarking can be at least partially theorised in terms of the larger dynamics of market competition between political causes and organisations for resources, audiences, allies, and credibility.

Much of the recent proliferation of global benchmarks can be traced to their perceived capacity to help build the reputation of specific organisations as

---

[16] Diego Giannone, 'Political and ideological aspects in the measurement of democracy: the Freedom House case', *Democratization*, 17:1 (2010), pp. 68–97.

[17] Antje Vetterlein, 'Seeing like the World Bank on poverty', *New Political Economy*, 17:1 (2012), pp. 35–58.

[18] Nehal Bhuta, 'Governmentalizing sovereignty: Indexes of state fragility and the calculability of political order', in Kevin E. Davis, Angelina Fisher, Benedict Kingsbury, and Sally Engle Merry (eds), *Governance by Indicators: Global Power Through Quantification* (Oxford: Oxford University Press, 2012), pp. 132–62.

[19] See the various contributions to Alexander Cooley and Jack Snyder (eds), *Ranking the World: Grading States as a Tool of Global Governance* (Cambridge: Cambridge University Press, 2015); see also Judith G. Kelley and Beth A. Simmons, 'Politics by number: Indicators as social pressure in International Relations', *American Journal of Political Science*, 59:1 (2015), pp. 55–70.

[20] See, for example, Thomas Risse, Stephen Ropp, and Kathryn Sikkink (eds), *The Power of Human Rights: International Norms and Domestic Change* (Cambridge: Cambridge University Press, 1999); Martha Finnemore and Kathryn Sikkink, 'International norm dynamics and political change', *International Organization*, 52:4 (1998), pp. 887–917; Rawi Abdelal, Mark Blyth, and Craig Parsons (eds), *Constructing the International Economy* (Ithaca, NY: Cornell University Press, 2010); Susan Park and Antje Vetterlein (eds), *Owning Development: Creating Policy Norms in the IMF and the World Bank* (Cambridge: Cambridge University Press); Alexander Betts and Phil Orchard (eds), *Implementation and World Politics: How International Norms Change Practices* (Oxford: Oxford University Press, 2014).

[21] See, for example, Charli Carpenter, 'Studying issue (non)-adoption in transnational advocacy networks', *International Organization*, 61:3 (2007), pp. 643–67; Charli Carpenter, 'Setting the advocacy agenda: Issues and non-issues around children and armed conflict', *International Studies Quarterly*, 51:1 (2007), pp. 99–120.

[22] See Charli Carpenter, *Lost Causes: Agenda-Setting and Agenda-Vetting in Global Issue Networks* (Ithaca, NY: Cornell University Press, 2014); Clifford Bob, *The Marketing of Rebellion: Insurgents, Media, and International Activism* (Cambridge: Cambridge University Press, 2005); Wendy Wong, *Internal Affairs: How the Structure of NGOs Transforms Human Rights* (Ithaca: Cornell University Press, 2012).

[23] Peter A. Gourevitch, David A. Lake, and Janice Gross Stein (eds), *The Credibility of Transnational NGOs: When Virtue is Not Enough* (Cambridge: Cambridge University Press, 2012); L. David Brown, *Creating Credibility: Legitimacy and Accountability for Transnational Civil Society* (London: Kumarian Press, 2008).

'issue experts'.[24] The popularity of benchmarking as a strategic tool for producing authoritative expertise – or at least the public appearance of expertise – is most notable in relation to NGOs and some IOs, which frequently find themselves in competition with their peers for allies, attention, and resources.[25] Thanks to the digital revolution of the last two decades,[26] it is often cheaper and easier to formulate and disseminate benchmarks than to engage in most forms of on the ground intervention. These conditions have contributed to an increasing level of market saturation, with NGOs, IOs, and other actors launching competing benchmarks as part of strategic efforts to create and consolidate a distinctive brand.

It is also important to take into account the intersections between expertise, authority, and indirect power. Over the last decade, a number of IR scholars have focused on the role of expert knowledge in the exercise of indirect power.[27] Recent works have demonstrated that expert knowledge and authority have helped to shape the architecture and practice of transnational governance,[28] to construct authority at the transnational level and effect transnational decision-making,[29] and to configure incentive systems that drive the global diffusion of common policy models and normative standards.[30]

The literature on rational design and institutional choice also offers further insights into benchmarking practices, most notably in relation to cooperation, coordination, and regulation. Rationalist theories can be loosely grouped together around the basic idea that both state and non-state actors often have a mutual interest in coordinating and codifying their activities across different spheres of global governance, and that these interests help to explain variations in the design and operation of international institutions. These overlapping interests may include shared efficiency gains, similar interests associated with information sharing and standardisation, and the mutual benefits gained from institutional arrangements that overcome collective action problems. This final point is based on the

---

[24] Peter J. May, Chris Koski, and Nicholas Stramp, 'Issue expertise in policymaking', *Journal of Public Policy*, DOI: 10.1017/S0143814X14000233; see also Leonard Seabrooke, 'Epistemic arbitrage: Transnational professional knowledge in action', *Journal of Professions and Organizations*, 1:1 (2014), pp. 49–64.

[25] Ellen Gutterman, 'The legitimacy of transnational NGOs: Lessons from the experience of transparency international in Germany and France', *Review of International Studies*, 40:2 (2014), pp. 391–418; see also Leonard Seabrooke and Duncan Wigan, 'How activists use benchmarks: Reformist and revolutionary benchmarks for global economic justice', *Review of International Studies*, 41:5 (2015), pp. 887–904. Ole Jacob Sending, *The Politics of Expertise: Competing for Authority in Global Governance* (Ann Arbor: University of Michigan Press, 2015), p. 12.

[26] Lance Bennett and Alexandra Segerberg, *The Logic of Connective Action: Digital Media and the Personalization of Contentious Politics* (Cambridge: Cambridge University Press, 2014).

[27] This has also been a major theme of work in cognate fields, such as Law. See Kevin E. Davis, Benedict Kingsburgy, and Sally Engle Merry, 'Introduction: the local-global life of indicators: Law, power, and resistance', in Sally Engle Merry, Kevin E. Davis, and Benedict Kingsbury (eds), *The Quiet Power of Indicators: Measuring Governance, Corruption, and Rule of Law* (Cambridge: Cambridge University Press, 2015), pp. 1–24.

[28] Oded Löwenheim, 'Examining the state: a Foucauldian perspective on international "governance indicators"', *Third World Quarterly*, 29:2 (2008), pp. 255–74; Jacqueline Best, *Governing Failure: Provisional Expertise and the Transformation of Global Development Finance* (Cambridge: Cambridge University Press, 2014); Michael Barnett and Martha Finnemore, *Rules for the World: International Organizations in Global Politics* (Ithaca, NY: Cornell University Press, 2004).

[29] Cornel Ban, *Ruling Ideas: How Global Economic Paradigms Go Local* (Oxford: Oxford University Press, 2016); André Broome and Leonard Seabrooke, 'Shaping policy curves: Cognitive authority in transnational capacity building', *Public Administration*, doi: 10.1111/padm.12179; Sending, *The Politics of Expertise*.

[30] J. C. Sharman, 'Power and discourse in policy diffusion: Anti-money laundering in developing states', *International Studies Quarterly*, 52:3 (2008), pp. 635–56; Kelley and Simmons, 'Politics by number'.

understanding that 'states and other international actors, acting for self-interested reasons, design institutions purposefully to advance their joint interests'.[31]

These types of arguments help to explain why a wide range of transnational actors have increasingly embraced benchmarks and benchmarking. As we explore in more detail below, much of the appeal of benchmarks stems from their capacity to translate complex phenomena into numerical information. This makes it feasible for non-experts to make comparisons across a diverse range of cases and contexts, and enables the definition of targets and numerical criteria that can facilitate evaluations of relative performance. From this vantage point, the recent proliferation of benchmarking can be at least partially traced to a combination of rational interests, market demand, and institutional design. This is most notable in relation to private market governance, where one of the main motivations behind benchmarking has been to produce useable information that improves how actors respond to market forces and conditions.

In addition to information sharing, benchmarks play an increasingly central role when it comes to standardising and coordinating corporate policies on issues such as labour conditions and corporate social responsibility.[32] Similarly, benchmarking efforts now play a key role in policy coordination and institutional design among states and IOs faced with collective action problems over climate change, disaster management, and human development.[33] These efforts have not always been successful in bringing about desired outcomes (and there are circumstances when benchmarking can be used to deflect pressure for larger reforms), but it is nonetheless clear that there are a number of occasions when benchmarking can be theorised as a product of rational interests and cooperation.

While the existing IR scholarship on human rights and international institutions offers useful insights, these are not sufficient enough to fully understand the practices and politics of global benchmarking. We therefore draw inspiration from a growing literature concerned with governmentality, the exercise of indirect power, and related technologies of rule over distant entities in the international arena.[34] Governmentality has proved especially popular within IR circles as a way to theorise how forms of liberal or neoliberal governance have been able to exercise power at a distance by both constraining and channelling the social, political, and institutional horizons of specific actors and institutions.[35] One of the key points at issue here is 'how certain identities

---

[31] Barbara Koremenos, Charles Lipson, and Duncan Snidal, 'The rational design of international institutions', *International Organization*, 55:4 (2001), p. 781; see also Joseph Jupille, Walter Mattli, and Duncan Snidal, *Institutional Choice and Global Commerce* (Cambridge: Cambridge University Press, 2013).

[32] See LeBaron and Lister, 'Benchmarking global supply chains'; James Harrison and Sharifah Sekalala, 'Addressing the compliance gap? UN initiatives to benchmark the human rights performance of states and corporations', *Review of International Studies*, 41:5 (2015), pp. 925–45.

[33] See Tony Porter, 'Global benchmarking networks: the cases of disaster risk reduction and supply chains', *Review of International Studies*, 41:5 (2015); Liam Clegg, 'Benchmarking and blame games: Exploring the contestation of the Millennium Development Goals', *Review of International Studies*, 41:5 (2015); Caroline Kuzemko, 'Climate change benchmarking: Constructing a sustainable future?', *Review of International Studies*, 41:5 (2015), all in this Special Issue.

[34] Carl Death, 'Governmentality at the limits of the international: African politics and Foucauldian theory', *Review of International Studies*, 39:3 (2013), p. 768; see also Ole Jacob Sending and Iver B. Neumann, 'Governance to governmentality: Analyzing NGOs, states, and power', *International Studies Quarterly*, 50:3 (2006), pp. 651–72.

[35] Governmentality approaches are based on the work of Michel Foucault, see Fougner, 'Neoliberal governance of states'; Jonathan Joseph, 'The limits of governmentality: Social theory and the international', *European Journal of International Relations*, 16:2 (2010), pp. 223–46; Wanda Vrasti, 'Universal but not truly "global": Governmentality, economic liberalism, and the international', *Review of International Studies*, 39:1 (2013), pp. 49–69.

and action-orientations are defined as appropriate and normal and how relations of power are implicated in these processes'.[36] This theme has particular resonance in the case of global benchmarking, because benchmarks primarily operate by quantifying and projecting normative criteria regarding the parameters of appropriate conduct and performance.[37]

The literature on governmentality is especially useful for the insight that benchmarking functions to make diverse forms of behaviour legible and amenable to intervention.[38] However, existing applications of governmentality, which mostly analyse techniques of government in domestic contexts, cannot simply be stapled on to analyses of the transnational arena. Among other things, transnational governance initiatives are characterised by a high degree of variation in both the rate and the form of implementation across different jurisdictions.[39] Governmentality approaches are also less useful for understanding how benchmarking practices facilitate the political agendas of specific actors and organisations. We suggest that global benchmarks can nonetheless be usefully located within larger patterns of governmentality associated with contemporary transnational governance. These patterns are closely associated with indirect forms of power that establish appropriate standards of behaviour across a wide variety of policy domains.[40] Individual benchmarks tend to overlap in multiple ways, and therefore contribute to the diffusion of normative visions and agendas regarding what transnational actors should look like, what they should value, and how they should behave.

Translating normative values into numerical representations

Global benchmarking tends to be heavily reliant upon rhetorical appeals to authoritative expertise. Instead of relying upon forms of direct compulsion (actor *A* compels actor *B* to do what *A* wants),[41] global benchmarks usually operate by orienting how specific actors: (1) conceptualise their options, obligations, and opportunities; and (2) seek to legitimate and justify their performance and perceived relative standing. It is in within this context that benchmarking practices can be regarded as an exercise in governance at a distance, which combines indirect power, expert authority, and transnational governmentality. This also means that the political effects of benchmarking tend to be cumulative and subtle, rather than overt and immediate, but they can nonetheless have a major influence over processes of agenda-setting in transnational governance.

The recent proliferation of global benchmarks owes a major debt to the political and popular appeal of numbers as information shortcuts, whereby complex and

---

[36] Iver B. Neumann and Ole Jacob Sending, *Governing the Global Polity: Practice, Mentality, Rationality* (Ann Arbor: University of Michigan Press, 2010), p. 10; see also Mitchell Dean, *Governmentality: Power and Rule in Modern Society* (Thousand Oaks, CA: Sage, 2009).

[37] See also Katja Freistein, 'Effects of indicator use: a comparison of poverty measuring instruments at the World Bank', *Journal of Comparative Policy Analysis: Research and Practice*, DOI:10.1080/13876988.2015.1023053.

[38] See, for example, Larner and Le Heron, 'Global benchmarking'; Fougner, 'Neoliberal governance of states'.

[39] Joseph, 'The limits of governmentality', p. 243.

[40] Neumann and Sending, *Governing the Global Polity*, p. 66; Lorenzo Fioramonti, *How Numbers Rule the World: The Use and Abuse of Statistics in Global Politics* (London: Zed Books, 2014); Sending, *The Politics of Expertise*.

[41] Michael Barnett and Raymond Duvall, 'Power in international politics', *International Organization*, 59:1 (2005), pp. 39–75.

contested normative values are translated into simplified numerical representations. This process of translation not only helps to obscure their normative foundations, it also enables non-experts to make crude comparisons of relative performance regarding complex phenomena at a transnational level. This translation process is common to all forms of benchmarking, and can be divided into four distinct components:[42]

- simplification and extrapolation
- commensuration
- reification
- symbolic judgement

Simplification and extrapolation are preconditions of quantification. Simplification comes in many different forms, but the most common denominator is when complexity and contextual detail is 'lost in translation' in the pursuit of quantification and comparability. Since not every sphere of human activity can be easily quantified, benchmarking efforts have a tendency to gravitate towards behaviours that can be more easily and effectively translated into a numerical form, and thereby end up generating data that is chiefly based upon a narrow subset of contributing factors. Simplification also tends to overlook context-specific idiosyncrasies and histories in favour of an emphasis upon more general properties. The inherent limitations of simplification are often further complicated by extrapolation, which refers to efforts to 'plug the gaps' when available data falls short. Quantification requires reliable and comprehensive information, yet reliable information can often be in short supply in many contexts and countries. Faced with persistent and significant shortfalls, benchmarkers can end up extrapolating based upon what they already know – or what they think they know – which can result in highly speculative findings that later take on the imprimatur of facts once they are translated into numerical form.

Another component of numerical translation is commensuration, which refers to 'the expression or measurement of characteristics normally represented by different units according to a common metric'.[43] Otherwise dissimilar political, economic, and social conditions become easily comparable by translating qualities into quantities. Once qualities are translated into quantities, they can be graded and assessed in terms of their orders of magnitude. Commensuration therefore imposes a form of homogeneity among disparate entities that is imagined to be 'a property of the object rather than something produced by quantification'.[44] In addition, there are further advantages associated with the 'neutrality' and 'objectivity' commonly ascribed to numerical rankings and representations. 'Numbers are not like words, which require interpretation', but are instead widely perceived to present unbiased facts.[45] There is a widespread tendency to fixate on specific numerical claims, which create 'anchoring effects' by establishing referents that shape how people later conceptualise specific issues.[46] These 'anchoring effects' also underpin the capacity of

---

[42] Simplification and extrapolation, commensuration, reification, and symbolic judgement are distinct components of the larger process of global benchmarking, rather than sequential phases.

[43] Wendy Nelson Espeland and Mitchell L. Stevens, 'Commensuration as a social process', *Annual Review of Sociology*, 24 (1998), p. 315.

[44] Ibid., p. 316.

[45] Fioramonti, *How Numbers Rule the World*, p. 192.

[46] Peter Andreas and Kelly M. Greenhill, 'Introduction: the politics of numbers', in Peter Andreas and Kelly M. Greenhill (eds), *Sex, Drugs, and Body Counts: The Politics of Numbers in Global Crime and Conflict* (Ithaca, NY: Cornell University Press, 2010), p. 17.

numbers to generate information in a format that can be more easily and quickly assimilated by non-expert audiences, who might otherwise be overwhelmed by qualitative and contextual detail.

Commensuration requires fixed, stable, and universal categories. These are generated by means of reification, which refers to the translation of complex phenomena into observable and quantifiable conceptual categories that are presumed to be universally applicable irrespective of cultural or historical context. Reification effectively stabilises the meaning of complex and highly contested categories, such as democracy, freedom, and stability.[47] These reified categories in turn provide a foundation for different types of numerical assessment, the most notable of which are rankings, scales, and grades. Ranking consists of assigning individual units of analysis a position relative to their peers, such as country A being number one while country B is number six. Numerical scales, such as 1–10, produce the appearance of more precise and fine-grained measurement, with the performance of different actors being assigned a specific score out of a fixed total. In contrast, grades classify and group together multiple peers into defined qualitative bands, such as Free to Not Free, or Tier One to Tier Three. Grades are frequently represented using 'heat maps', with shades of green, yellow, and red being assigned to countries on a regional or global map based upon the specific grades they have been awarded. As benchmarking systems have evolved, the types of numerical assessments they generate have become ever more elaborate, but none of these assessments are possible without a foundation of fixed and unproblematic categories that create the appearance of certainty, coherence, and consistency.

Quantification and reification pave the way for symbolic judgments, in which the question of relative performance or value takes centre stage. Symbolic judgements on countries' relative performance are qualitatively different from what can be termed 'regulatory judgements', such as a determination that a government's actions constitute non-compliance with prescribed behaviour under the terms of an international agreement.[48] *Regulatory judgements* are more likely to involve direct and easily observable political consequences on target actors, whereas *symbolic judgements* are more likely to produce indirect political consequences through shaming processes, unfavourable comparisons with peers, and other forms of reputational damage. They may also generate reference points that are carried into other types of transnational practices, such as multilateral lending and development assistance, bilateral diplomatic relations, access to capital markets, or international programmes for intervention and policy reform.

Nearly all global benchmarks suffer from a 'dodgy data' problem. This problem can be particularly acute in cases where many different benchmarks are used in order to create composite benchmarks, resulting in a proliferation of data that frequently rests on very tenuous foundations.[49] Many benchmarkers are reluctant to make their methodology public, since this could complicate or undercut their market position or organisational credibility. Therefore, it often remains a mystery how specific

---

[47] Theodore M. Porter, *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life* (Princeton: Princeton University Press, 1995), p. 86.
[48] Beth A. Simmons, 'Compliance with international agreements', *Annual Review of Political Science*, 1 (1998), p. 77.
[49] Kevin E. Davis, Benedict Kingsbury, and Sally Engle Merry, 'Indicators as a technology of global governance', *Law and Society Review*, 46:1 (2012), p. 76.

conclusions were reached. This backstory routinely gets obscured once numbers are put into the public domain.

In many cases, the main political and institutional advantages associated with creating and disseminating benchmarks are political and organisational, rather than analytical. Global benchmarks have not only become relatively cheap and easy to produce and disseminate, they have also become increasingly popular among funders and donors eager to capture media attention. Most benchmarking does not involve years of expertise in the field, or contextual knowledge of local languages, customs, and social norms. Instead, all that is often required is the capacity to compile and process different forms of secondary data, which may simply involve aggregating and transposing information from one benchmark in order to create another.

The politics of global benchmarking

Global benchmarking typically relies on productive forms of indirect power to provoke reactions from target actors, with 'productive power' understood as the 'socially diffuse production of subjectivity in systems of meaning and signification'.[50] Much of the value of benchmarking, at least from a public relations or political activism standpoint, stems from the fact that benchmarks can play a key role in both stimulating and structuring political conversations regarding: (1) the dimensions, ramifications, and salience of a given set of issues; (2) how the performance of specific actors compares with that of their peers; and (3) how the performance of specific actors has changed with the passage of time. Benchmarking practices also tend to provoke politically motivated conversations around questions of methodology, whereby the credibility of particular measures is either impugned or defended depending on whether the results align with the political and economic agendas of the various actors involved.

While benchmarks purport to describe 'things as they are', this veneer of numerical representation and neutral comparison invariably conceals a range of political calculations, agendas, interests, and effects. Any overall assessment of 'good' or 'bad' performance requires a series of prior normative judgements regarding the types of activities, institutions, or categories that merit being subjected to benchmarking in the first place. At this juncture, it is essential to recognise that global benchmarking efforts almost invariably draw upon a common portfolio of normative values, assumptions, and agendas, such as liberal or neoliberal models of the rule of law, freedom of speech, democracy, human development, environmental protection, poverty alleviation, 'modern' statehood, and 'free' markets.

These normative commitments typically have a similar point of origin and influence, with Western experiences, assumptions, and paradigms exercising a disproportionate influence over the shape of international policy agendas and the articulation and definition of global problems. Moreover, Western states tend to populate the highest rankings across numerous benchmarks, with many non-Western states in turn receiving the lowest scores. While this is by no means a perfect relationship, since some non-Western states now feature amongst the 'high achievers', it is still possible to identify clear concentrations of Western and non-Western states at opposing ends of the spectrum across many global benchmarks.

---

[50] Barnett and Duvall, 'Power in international politics', p. 43.

	HDI*	Corruption	Freedom	Stability	Credit	Slavery	Business
High-performing European countries							
Norway	1	5	1 (free)	175	Aaa	150	9
Switzerland	3	7	1 (free)	174	Aaa	150	29
Netherlands	4	8	1 (free)	166	Aaa	139	28
Germany	6	12	1 (free)	165	Aaa	136	21
Denmark	10	1	1 (free)	176	Aaa	150	5
Ireland	11	21	1 (free)	170	Baa1	160	15
Sweden	12	3	1 (free)	177	Aaa	150	14
Iceland	13	12	1 (free)	171	Baa3	160	13
UK	14	14	1 (free)	161	Aa1	160	10
France	20	22	1 (free)	160	Aa1	139	38
Low-performing African countries							
Mozambique	178	119	3.5 (partly free)	50	B1	35	139
Guinea	179	150	5 (partly free)	12	Unrated	17	175
Burundi	180	157	5 (partly free)	21	Unrated	41	140
Burkina Faso	181	83	4 (partly free)	39	Unrated	26	154
Eritrea	182	160	7 (not free)	23	Unrated	15	184
Sierra Leone	183	119	3 (partly free)	35	Unrated	13	142
Chad	184	163	6.5 (not free)	6	Unrated	34	189
CAR**	185	144	7 (not free)	3	Unrated	31	188
Congo (DRC)	186	154	6 (not free)	4	B3	23	183
Niger	187	106	3.5 (partly free)	19	Unrated	28	176
N =	187	177	195	178		162	189
Metric	Ranking (lower better)	Ranking (lower better)	(lower better)	Ranking (higher better)	Grade (higher better)	Ranking (higher better)	Ranking (lower better)

Notes: *The table is composed of the 10 best-performing European and 10 worst-performing African states which feature across multiple benchmarks based on the 2014 Human Development Index (HDI). **Central African Republic.

Sources: 2014 Human Development Report; 2013 Corruption Perceptions Index; 2014 Freedom in the World Report; 2014 Fragile State Index; Moody's Investors Service (accessed 9 October 2014); 2013 Global Slavery Index; 2013 World Bank Doing Business Survey.

Table 1. *Comparing European and African countries across global benchmarks*

Table 1 serves to illustrate this underlying relationship by comparing the rankings recently assigned to ten 'high performing' European and ten 'low performing' African countries across a number of high-profile global benchmarks, including human development, corruption, freedom, state stability, credit, slavery, and business. The selection of ten countries is deliberate, since press releases and other materials that accompany the publication of benchmarks frequently concentrate attention on the top ten 'best' or 'worst' performers, who are specifically singled out for either condemnation or praise.

Each benchmark listed in this table is global in both scope and ambition. This global reach builds upon an underlying premise that there are certain values and criteria that can and should be treated as universal, irrespective of historical, political, or cultural differences. This type of universalism generates considerable controversy and norm contestation when expressed in other formats,[51] but benchmarks have proved to be an effective means of at least partially shielding normative arguments and agendas via appeals to models of neutral and technical assessment. Since the underlying normative commitments associated with an individual benchmark often closely align with those of other benchmarks in related domains, global benchmarking tends to have the cumulative effect of: (1) both reifying and generalising specific

[51] Kelley, *Monitoring Democracy*, p. 23.

models of governance, social organisation, and public policy; and (2) legitimating and promoting the recent histories and ongoing activities of Western states and, by extension, a variety of Western transnational actors.

In our opening discussion we divided benchmarking into three areas: quality of conduct, quality of design, and quality of outcomes. In the case of quality of outcomes, it is important to take into account a widespread tendency to assign singular responsibility for 'good' or 'bad' outcomes to the internal efforts of states and their peoples. When the United Kingdom receives a positive ranking, it is presumed to be the result of the internal efforts of the British state and its citizens, rather than as a consequence of interactions between Britain and other parts of the globe. Similarly, when Nigeria receives a negative ranking, it is tacitly presumed to be a result of the internal failings of the Nigerian state and society, rather than a consequence of external intrusions or structural conditions in the international system. This is highly problematic from an analytical standpoint, because the sources of 'good' or 'bad' performance tend to be far more diffuse than this model of responsibility suggests. There are many occasions when 'successful' states, along with numerous non-state actors, are at least partially responsible for the 'failures' of their peers. To give a stark example that illustrates this point: Iraq today scores poorly on a host of benchmarks, but how much of this is the responsibility of Iraqis?

This analytical slippage between outcomes and responsibility can be politically valuable for Western governments, populations, and corporations. Since high scores are widely presumed to be the result of individual efforts and achievements, global benchmarks frequently end up tacitly legitimating the wealth and privilege enjoyed by many actors in the West. Since low scores are widely presumed to be the result of internal failings and shortcomings, the impact of external actors and forces – most notably colonialism and imperialism – gets excluded from the political calculus.[52] This basic formula is in turn likely to provide further justification for particular forms of intervention and analysis, whereby Western actors can be represented as saviours and non-Western actors can be reduced to supplicants in need of paternalistic assistance. This formula obviously comes with a host of problems. In particular, no benchmark that assigns responsibility for outcomes that it is beyond the capacity of the 'responsible' party to address will be effective in bringing about change.[53]

These languages of legitimation and exculpation comprise one component of the larger politics of 'good' and 'bad' performance. In the case of the former, benchmarks tend to help reinforce established policies and organisational practices through the validating effects of favourable scores and superior rankings. This also extends to improvements in performance, where political leaders and other actors routinely claim credit when their countries and organisations have improved in global rankings, and may also seek to harness 'improvement' in order to attract interest from investors and aid agencies.[54] In these types of cases, benchmarks frequently become an

---

[52] This inward attribution of Western success is explored further in John Hobson, *The Eastern Origins of Western Civilisation* (Cambridge: Cambridge University Press, 2004); see also Shogo Suzuki, *Civilization and Empire: China and Japan's Encounter with European International Society* (Abingdon: Routledge, 2009); on the assessment of 'legitimacy' in terms of Western policy standards, see Lena Rethel, 'Whose legitimacy? Islamic finance and the global financial order', *Review of International Political Economy*, 18:5 (2011), pp. 75–98.

[53] See Ole Jacob Sending and Jon Harald Sanders Lie, 'The limits of global authority: How the World Bank benchmarks economies in Ethiopia and Malawi', *Review of International Studies*, 41:5 (2015), pp. 993–1010.

[54] Alexander Cooley, 'The emerging politics of international ranks and ratings: a framework for analysis', in Alexander Cooley and Jack Snyder (eds), *Ranking the World: Grading States as a Tool of Global Governance* (Cambridge: Cambridge University Press, 2015), p. 4.

instrument of status quo legitimation, and may be further invoked in order to deflect or dismiss calls for a different course of action.

The politics of 'bad' performance pull in a different direction, with negative or falling rankings providing an impetus for overhauling existing laws and policies, or at least providing political ammunition for critics of the status quo. Here, benchmarks can potentially prompt actors to 'alter their behaviour in reaction to being evaluated, observed, or measured'.[55] This can occur either *ex ante*, when actors anticipate future costs associated with a benchmarking exercise and seek to avoid the possibility of reputational damage, or *ex post*, when target actors observe and then respond to the costs associated with a specific result.[56] Unfavourable rankings in different global benchmarking regimes may result in either material sanctions (such as economic costs) or social sanctions (such as shaming or peer pressure via instruments such as a 'watch list' or a 'blacklist'), or a mix of both.[57]

There are many instances where a 'poor' result may have little or no immediate political effects; neither the material nor the social sanctions associated with benchmarking have consistent or predictable effects upon the behaviour of target actors. The imposition of material sanctions on 'pariah' states has often proved to be counterproductive for altering behaviour,[58] while those that have already gained pariah status are unlikely to be constrained by being further shamed and ostracised by the international community or through other social sanctions.[59] Nonetheless, when benchmarks gain sufficient prominence and credibility to provide a strong rationale for political action, they can exert a significant influence as a means to 'legitimate policy goals, the choice of target populations, and policy tools'.[60]

The degree of analytical and methodological rigour that underpins the construction of global benchmarking regimes cannot sufficiently explain why they have emerged as such a popular mode of transnational governance. A more compelling explanation is that the growth of global benchmarking reflects a dynamic 'benchmarking market'. This is tied to growing demand for benchmarks as a form of 'evidence' to enhance broader processes of governance, such as the effective allocation of official development assistance, the identification of internal security threats, enhancing accountability mechanisms in transnational governance, tracking standards of corporate behaviour, or monitoring national compliance with international policy regimes. There will therefore be occasions when 'the demand for numbers generates a supply'.[61] Yet while rank orderings of conduct, institutional design, and economic, social, and political outcomes may fulfil a functional need for

---

[55] Wendy Nelson Espeland and Michael Sauder, 'Rankings and reactivity: How public measures recreate social worlds', *American Journal of Sociology*, 113:1 (2007), p. 6.

[56] J. C. Sharman, 'The bark *is* the bite: International organizations and blacklisting', *Review of International Political Economy*, 16:4 (2009), pp. 573–96.

[57] J. C. Sharman, *Havens in a Storm: The Struggle for Global Tax Regulation* (Ithaca, NY: Cornell University Press, 2006), p. 104.

[58] Alexandra Homolar, 'Rebels without a conscience: the evolution of the rogue states narrative in US security policy', *European Journal of International Relations*, 14:4 (2010), pp. 705–27.

[59] Edward Weisband, 'Discursive multilateralism: Global benchmarks, shame, and learning in the ILO labor standards monitoring regime', *International Studies Quarterly*, 44:4 (2000), pp. 643–66.

[60] Anne Schneider and Helen Ingram, 'Social construction of target populations: Implications for politics and policy', *American Political Science Review*, 87:2 (1993), p. 339.

[61] Peter Reuter and Edwin M. Truman, *Chasing Dirty Money: The Fight Against Money Laundering* (Washington, DC: International Institute of Economics, 2004), p. 22, available at: {www.piie.com/publications/chapters_preview/381/2iie3705.pdf} accessed 20 June 2014.

existing processes of transnational governance, they also produce new power relations wielded by one group of actors over others.[62]

The practice of global benchmarking is a prime example of transnational governance that works via knowledge practices rooted in authoritative expertise in order to extend power over disparate objects and subjects.[63] However, benchmarking is distinct from other forms of expert authority commonly utilised by state institutions and international organisations, because of the opportunities it provides for non-state actors – whether civil society organisations or corporate agencies – to employ knowledge practices in an attempt to limit or alter how public authority is used. It is therefore important to unpack the practice of global benchmarking into different types to gain a more fine-grained understanding of how various forms of benchmarking, promulgated by different types of actors, intersect, overlap, and compete with each other across contemporary processes of transnational governance.

A typology of global benchmarking

In Table 2, we distinguish between four types of global benchmarking practices: (1) statecraft; (2) international governance; (3) private market governance; and (4) transnational advocacy.[64] This divides benchmarking practices into types based on the class of actor that is engaged in benchmarking, namely states, international organisations, profit-based private institutions, and non-profit private institutions. We use the public-private distinction as a 'category of analysis' to denote the different forms of accountability and capacities of various benchmarkers, rather than as a 'category of practice'.[65] While useful for heuristic purposes, these analytic divisions do not preclude the possibility that other actors may use one type of global benchmarking for a different purpose. Using this typology, we have compiled a Global Benchmarking Database consisting of 205 benchmarks (as of June 2015), which is available at: {www.warwick.ac.uk/globalbenchmarking/database}.

Type I benchmarking is a form of *statecraft*, whereby global benchmarks are produced by national government agencies such as ministries of finance and foreign affairs to extend state power internationally through the projection of particularistic values and standards of behaviour as universal. This may also legitimate the use of other foreign policy tools, such as sanctions and foreign aid, based on the conception of benchmark judgements as objective and neutral assessments of conduct, institutional design, or performance. Type II benchmarking is a form of *international governance*, which is undertaken by international organisations such as the World Bank, the International Monetary Fund (IMF), the Organisation for

---

[62] Tim Büthe, 'Beyond supply and demand: a politico-economic conceptual model', in Kevin E. Davis, Angelina Fisher, Benedict Kingsbury, and Sally Engle Merry (eds), *Governance by Indicators: Global Power Through Quantification* (Oxford: Oxford University Press, 2012), p. 51; see also Joel Quirk, 'The anti-slavery project: Linking the historical and the contemporary', *Human Rights Quarterly*, 28:3 (2006), p. 576.

[63] Wendy Larner and William Walters, 'Globalization as governmentality', *Alternatives*, 29:5 (2004), p. 496; see also André Broome and Leonard Seabrooke, 'Seeing like an international organisation', *New Political Economy*, 17:1 (2012), pp. 1–16.

[64] On the construction of typologies see Alexander L. George and Andrew Bennett, *Case Studies and Theory Development in the Social Sciences* (Cambridge: MIT Press, 2005), pp. 237–8.

[65] Stein Sundstøl Eriksen and Ole Jacob Sending, 'There is no global public: the idea of the public and the legitimation of governance', *International Theory*, 5:2 (2013), pp. 213–37.

Types of global benchmarking	Monitoring agents	Examples
Statecraft	National government agencies	Trafficking in Persons Report Millennium Challenge Account 　Country Rankings World Military Expenditures and Arms 　Transfers Rankings
International governance	International organisations and regional institutions	Human Development Index 　Gender Empowerment Measure 　World Governance Indicators 　Country Policy and Institutional 　　Assessment
Private market governance	Profit-based institutions, including financial services and consultancy firms	Emerging Markets Bond Indices 　International Country Risk Guide 　Sovereign Credit Ratings 　Supply chain benchmarking
Transnational advocacy	Civil society organizations, think tanks, media organisations, and academics	Corruption Perception Index 　Fragile/Failed States Index 　Index of Economic Freedom 　Climate Change Performance Index

Table 2. *Four types of global benchmarking practices*

Economic Cooperation and Development, and the United Nations Development Programme; or by regional organisations, such as the European Union. This differs from benchmarking as statecraft because the practice of Type II benchmarking is usually under the control of international bureaucracies rather than national policymakers and is less directly geared towards the promotion of an individual state's national interests, although states often seek to use Type II benchmarks as instruments of statecraft.

　　Type III benchmarking is a form of *private market governance*, which is undertaken by profit-based institutions and is one of the oldest forms of benchmarking. This includes sovereign credit rating, which has its roots in the late nineteenth and early twentieth centuries,[66] and internal measures of performance and quality control used by large firms (self-benchmarking),[67] which has become increasingly significant as transnational corporations have spread their business activities worldwide through global production networks. Type IV benchmarking is either explicitly or implicitly geared towards *transnational advocacy* in particular issue areas, and is primarily conducted by civil society organisations and non-profit think tanks, but may also include work by individual academics or academic research centres. In some instances Type IV benchmarking involves collaboration between non-profit institutions and profit-based institutions, and in particular media organisations, such as the Index of Economic Freedom, which is produced by the Heritage Foundation and the *Wall Street Journal*. We further illustrate our typology of global benchmarking practices by briefly discussing a prominent example of each of type.

[66] Timothy J. Sinclair, *The New Masters of Capital: American Bond Rating Agencies and the Politics of Creditworthiness* (Ithaca, NY: Cornell University Press, 2005).
[67] Larner and Le Haron, 'Global benchmarking'.

Benchmarking as statecraft

Benchmarking as statecraft can be conceived as a form of 'soft power' in world politics.[68] A prominent example of Type I benchmarking is the *Trafficking in Persons Report*, which has been produced annually since 2001 by the US State Department's Office to Monitor and Combat Trafficking in Persons, and is officially described as 'the U.S. Government's principal diplomatic tool to engage foreign governments on human trafficking'.[69] This was established through the Victims of Trafficking and Violence Prevention Act, signed into law in 2000, which mandates that unilateral sanctions be applied by the US government on countries that are deemed not to meet minimum standards for the elimination of human trafficking, based on the *Report*. These sanctions can involve exclusion from US non-humanitarian and non-trade-related foreign aid, as well as US opposition to government requests for IMF or World Bank loan programmes.

The *Report* divides countries based on three different tiers. Tier One comprises countries whose governments are assessed as fully compliant with minimum standards for the elimination of human trafficking. In Tier Two are countries whose governments are not fully compliant but are assessed as making significant efforts to comply, with those deemed to face severe problems included in a separate category on the Tier Two Watch List. In Tier Three are countries whose governments are judged as non-compliant and not making sufficient efforts to comply with these minimum standards.[70] The data for the *Report* is based upon information from US embassies, government officials, non-governmental and international organisations, published reports, news articles, academic studies, research trips to every region of the world, and information submitted via email to report tips on human trafficking.[71]

In the decade and a half since it was established, the *Trafficking in Persons Report* has attracted substantial controversy and has been criticised for a lack of impartiality and political bias, with the US accused of acting as a 'global sheriff'.[72] States that have posed significant foreign policy problems for the US, such as Cuba and Venezuela, have typically received poorer rankings than otherwise broadly similar countries. Allies of the US with questionable records on human rights have historically received more positive assessments, although there has recently been a modest effort to correct this perception by shaming some US allies.[73] The US government has also recently found it necessary to include material on its own anti-trafficking efforts, following sustained criticism that their own record was notably absent from the reports. While all benchmarkers invariably start with specific agendas of their own, it is not uncommon for annual benchmarking

---

[68] Joseph S. Nye, *Soft Power: The Means to Success in World Politics* (New York: Public Affairs, 2004).
[69] {http://www.state.gov/j/tip/rls/tiprpt/} accessed 11 July 2014.
[70] Anne Gallagher, 'Trafficking in Persons Report (Review)', *Human Rights Quarterly*, 23:4 (2001), pp. 1136–7; Anne Gallagher, 'Improving the effectiveness of the international law of human trafficking: a vision for the future of the US Trafficking in Persons Reports', *Human Rights Review*, 12:3 (2011), pp. 381–400.
[71] {http://www.state.gov/j/tip/rls/tiprpt/2014/226645.htm} accessed 11 July 2014.
[72] Janie Chuang, 'The United States as global sheriff: Using unilateral sanctions to combat human trafficking', *Michigan Journal of International Law*, 27:2 (2006), pp. 437–94; Annie Bunting and Joel Quirk, 'Contemporary slavery as more than rhetorical strategy', in Annie Bunting and Joel Quirk, *The Invention of Contemporary Slavery: Studies in Rhetoric and Practice* (Vancouver: University of British Columbia Press, 2016).
[73] Gallagher, 'Improving the effectiveness of the international law of human trafficking', pp. 382–4.

exercises to evolve in unexpected ways, or to produce unpredictable findings and outcomes that complicate the original motivations for introducing the benchmark.

The *Report* is widely recognised as a prime example of the use of benchmarking as an exercise in statecraft that seeks to compel global action in accordance with the expectations and agenda of the US government.[74] Despite the numerous flaws that have been identified in these annual reports and associated policies responses,[75] recent research by Judith Kelley and Beth Simmons suggests that this example of Type I benchmarking has been highly consequential for the behaviour of (some) target actors. Kelley and Simmons conclude that 'states are sensitive to monitoring, respond faster to "harsher" grades, and react when their grade first drops below a socially significant threshold'.[76] Combating trafficking is a cause that comes with a host of practical problems and collateral damages, and it remains an open question whether the *Report* has helped or hurt in this respect. Nevertheless, this does not negate the larger point that this is a benchmark that has been globally influential.[77]

Benchmarking as international governance

The growth of benchmarking as international governance has gone hand in hand with the expansion of various forms of surveillance by international organisations of country performance over the last four decades.[78] The World Bank's Worldwide Governance Indicators (WGIs) is a useful illustrative example of Type II benchmarking. Starting from 1996 and covering over 200 countries, these indicators aim to measure governance performance around the world across six dimensions: (1) voice and accountability; (2) political stability and absence of violence; (3) governance effectiveness; (4) regulatory quality; (5) rule of law; and (6) control of corruption.[79] The data for the WGIs incorporates several hundred variables from 31 different data sources and is based on perceptions of governance quality drawn from public opinion and expert surveys, civil society organisations, profit-based information providers, and government agencies.

The conceptual validity, data accuracy, and substantive meaning of the WGI measures have been subject to strong criticism.[80] For example, to construct global benchmarks of governance quality, governance is defined as:

the traditions and institutions by which authority in a country is exercised. This includes (a) the process by which governments are selected, monitored, and replaced; (b) the capacity of

---

[74] Chuang, 'The United States as global sheriff', pp. 456–7.

[75] Ronald Weitzer, 'New directions in research on human trafficking', *The ANNALS of the American Academy of Political and Social Science*, 653:1 (2014), pp. 6–24; Thomas Steinfatt, 'Sex trafficking in Cambodia: Fabricated numbers versus empirical evidence', *Crime, Law, and Social Change,* 56:5 (2011), pp. 443–62.

[76] Kelley and Simmons, 'Politics by number', p. 68.

[77] See, for example, the articles in 'Beyond Trafficking and Slavery', *openDemocracy*, available at: {https://www.opendemocracy.net/beyondslavery} accessed 3 June 2015.

[78] See André Broome and Leonard Seabrooke, 'Seeing like the IMF: Institutional change in small open economies', *Review of International Political Economy*, 14:4 (2007), pp. 576–601; Liam Clegg, 'Our dream is a world full of poverty indicators: the US, the World Bank, and the power of numbers', *New Political Economy*, 15:4 (2010), pp. 473–92; Löwenheim, 'Examining the state'.

[79] Langbein and Knack, 'The worldwide governance indicators', pp. 369–70.

[80] Ibid., p. 351.

the government to effectively formulate and implement sound policies; and (c) the respect of citizens and the state for the institutions that govern economic and social interactions among them.[81]

Observers have highlighted the partial and biased view of governance quality that the definition used for the WGIs represents, emphasising in particular that the scale of aggregation involved in the production of the WGIs constitutes a trade-off between reliability and precision. Each of the data sources used to produce the WGIs suffers from specific quality problems. These problems are likely to be further complicated by aggregation processes, since the number and type of data sources differ both across countries and over time.[82]

The WGIs implicitly assume a particular meaning of governance as a universally accepted standard. As M. A. Thomas points out, while most governments are likely to agree that the 'rule of law' is an important dimension of effective governance, for a liberal democracy this might be understood as 'a state constrained by rules' while an authoritarian dictatorship might understand this to mean 'citizen obedience to government edicts'. For these reasons, the WGIs have been criticised for not recognising that 'a governance indicator is a hypothesis about measurement *and* about the nature of governance'.[83] Nevertheless, as an example of Type II benchmarking the WGIs have resonated across a wide range of third parties, and have become particularly influential in decision-making processes over foreign aid allocations as a new form of policy conditionality.[84]

Benchmarking as private market governance

Benchmarking has become an increasingly prominent feature of national and transnational economic governance, especially in the aftermath of the global financial crisis when a large proportion of the pre-crisis ratings of financial assets produced by credit rating agencies were found to be inaccurate.[85] In particular, sovereign credit ratings represent one of the most controversial examples of benchmarking as private market governance. Credit ratings are evaluations of a debtor's ability to repay a loan and the probability of default. As a form of Type III benchmarking, sovereign credit ratings by the three major ratings agencies – Moody's Investors Service, Standard and Poor's, and Fitch Ratings – impact upon governments' fiscal autonomy and the terms on which they can raise public debt. The symbolic judgements made by private ratings agencies affect the creditworthiness of national and local governments, firms, banks and other private companies, and in theory function to reduce uncertainty and information asymmetry problems for investors. Like other types of benchmarking,

---

[81] Aart Kraay, Daniel Kaufmann, and Massimo Mastruzzi, 'The worldwide governance indicators: Methodology and analytical issues', World Bank Policy Research Working Paper No. 5430 (2010), p. 4, emphasis in original, available at: {http://elibrary.worldbank.org/doi/pdf/10.1596/1813-9450-5430} accessed 15 July 2014.

[82] Steven van de Walle, 'The state of the world's bureaucracies', *Journal of Comparative Policy Analysis: Research and Practice*, 8:4 (2006), pp. 439–40.

[83] M. A. Thomas, 'What do the worldwide governance indicators measure?', *European Journal of Development Research*, 22:1 (2010), p. 50, emphasis added.

[84] Thomas, 'What do the worldwide governance indicators measure?', p. 32.

[85] Rawi Abdelal and Mark Blyth, 'Just who put you in charge? We did', in Alexander Cooley and Jack Snyder (eds), *Ranking the World: Grading States as a Tool of Global Governance* (Cambridge: Cambridge University Press, 2015), p. 46.

however, credit ratings 'not only provide information but help construct the context in which corporations and public bodies make decisions'.[86]

Standard and Poor's utilises a mix of qualitative and quantitative measures in the five factors that constitute its sovereign ratings. These include: (1) a 'political score', which focuses on the quality of political and policymaking institutions, and external risks; (2) an 'economic score', which incorporates the degree of economic diversity, income levels, and growth prospects; (3) an 'external score', based on the international status of a country's currency, external liquidity, and foreign debt levels; (4) a 'fiscal score', based on assessments of the sustainability of budget deficits and public debt burden; and (5) a 'monetary score', which is based on inflation rates, the degree of flexibility in monetary policy, and the depth of domestic financial markets. Standard and Poor's credit analysts assign a score for each of these five factors ranging from one (strongest) to six (weakest).[87] Once a sovereign credit rating is officially assigned to a country, ratings are then monitored on an ongoing basis and reviewed at least once a year.

Ratings issued by the three major agencies constitute a rank ordering of credit risk. Long-term ratings, for example, are distinguished between different ranks of 'investment grade' ratings (ranging from the top AAA rating, issued by Standard and Poor's and Fitch Ratings, to the BBB rating) and 'non-investment grade' or 'speculative grade' ratings (BB ratings and below.) Ratings below investment grade are considered to have a moderate to high credit risk of non-repayment. The power of the symbolic judgements rating agencies issue comes primarily from their role as 'reputational intermediaries'. This is based on their public image as independent, authoritative actors that are capable of making accurate expert assessments of creditworthiness, despite this image being subjected to stringent criticisms in recent years.[88] Moreover, credit rating agencies also remain highly controversial because their benchmarking activities help to reify and consolidate international norms of 'proper fiscal conduct', which shapes perceptions about what constitutes 'normal' economic behaviour by governments.[89]

Benchmarking as transnational advocacy

The use of global benchmarking by NGOs engaged in transnational advocacy has risen dramatically in recent years. Among other reasons for this growing trend, many funders value the capacity of benchmarks and indicators to provide 'clear signals' that can be used as proxies for measuring the relative success of political campaigns and policy interventions.[90] In June 2014, for example, the US think-tank Global Fund for Peace launched the tenth edition of their 'Fragile States Index', which is an example of Type IV benchmarking as transnational advocacy. Thanks to an established partnership with *Foreign Policy* magazine, the headline findings of the index had

---

[86] Fioramonti, *How Numbers Rule the World*, pp. 60–1.

[87] Standard and Poor's, *Sovereign Government Rating Methodology and Assumptions* (New York: Standard and Poor's, 2011) available at: {www.standardandpoors.com/ratings/articles/en/us/?articleType=PDF&assetID=124531655404} accessed 15 July 2014.

[88] Sinclair, *The New Masters of Capital*, p. 176.

[89] Bartholomew Paudyn, 'Credit rating agencies and the sovereign debt crisis: Performing the politics of creditworthiness through risk and uncertainty', *Review of International Political Economy*, 20:4 (2013), pp. 799–800.

[90] Sarah Bush, *The Taming of Democracy Assistance: Why Democracy Promotion Does Not Confront Dictators* (Cambridge: Cambridge University Press, 2015), p. 14.

been precirculated. Among the highlights of the 2014 report was the news that 'after six years in the number one position [as the world's worst 'failed state'] Somalia has finally been overtaken, leaving South Sudan as the most fragile state in the world'.[91]

The publication of the Fragile States Index (FSI) is intended as a form of transnational advocacy. With the stated goal of encouraging 'discussion, advocacy and action on the underlying conditions that could create conflict and do threaten human security and economic development',[92] the FSI ranks 178 states on the basis of their 'levels of stability and the pressures they face'.[93] This high-profile exercise in global benchmarking involves assigning states a numerical value (1–120) based on their relative vulnerability to 'state failure', and grouping states into different categories ranging from 'high alert' to 'very sustainable', with shades of red, yellow, and green used to highlight their relative status. These findings are generated using a patented 'Conflict Assessment Software Tool', which applies 'sophisticated search parameters and algorithms' to separate relevant from irrelevant data in the analysis of millions of documents each year.[94] This software tool integrates data from 'twelve primary social, economic and political indicators' (which include over 100 sub-indicators), with themes such as 'state legitimacy', 'factionalized elites', 'group grievance', 'security apparatus', and 'poverty and economic decline', yet the specific sources of raw data associated with these indicators have not been revealed.

The 'failed state' concept has been subject to sustained critique. Some of these criticisms were partly acknowledged by the decision by the Fund for Peace to rename the benchmark in 2014 as the 'Fragile States Index' to replace the former title of the 'Failed States Index'. The concept of 'failed states' has been repeatedly denounced for – amongst other things – lumping together states with very different histories and problems, for normalising a particular vision of 'modern' statehood and 'state-building', for directing responsibility for 'failure' inwards, rather than looking at external actors, and for being too closely aligned with US foreign policy goals.[95] This highlights the ease with which other actors might use one type of global benchmarking for different purposes, in order to pursue a larger set of interests and agendas.

Global benchmarking and third party users

The most influential users of each of these four types of global benchmarking are often third parties. These can be either public or private actors who may not be the formal target of a particular benchmarking exercise, but who incorporate benchmark scores produced by other actors into their decision-making processes and

---

[91] Fund for Peace, *Failed States Index 2014: Somalia Displaced as Most-Fragile State*, available at: {http://library.fundforpeace.org/fsi14-overview} accessed 15 July 2014.

[92] Fund for Peace, *Press Release: Fragile States Index 2014 Released*, available at: {http://library.fundforpeace.org/fsi14-pressrelease} accessed 3 June 2015.

[93] J. J. Messner (ed.), *Failed States Index 2014* (Washington: Global Fund for Peace, 2014), available at: {http://library.fundforpeace.org/library/cfsir1423-fragilestatesindex2014-06d.pdf} accessed 10 July 2014, p. 3.

[94] Messner, *Failed States Index*, p. 9.

[95] Charles Call, 'The fallacy of the "failed state"', *Third World Quarterly*, 29:8 (2008), pp. 1491–507; Ruth Gordon, 'Saving failed states: Sometimes a neocolonialist notion', *American University International Law Review*, 12:6 (1997), pp. 903–74; Olivier Nay, 'International organisations and the production of hegemonic knowledge: How the World Bank and the OECD helped invent the fragile state concept', *Third World Quarterly*, 35:2 (2014), pp. 210–31.

advocacy efforts.[96] This use by third parties can greatly expand the political traction of benchmarks by multiplying the reputational costs or benefits associated with specific rankings, and intensifying competitive pressures to improve poor performance.

Freedom House, for example, does not carry much independent weight as an organisation. As an advocacy-oriented NGO, it is unable to use material incentives to induce compliance; its symbolic judgements on country performance do not imply the same potential for direct consequences as negative country reports issued by international organisations such as the World Bank or the IMF, and its claim to expert authority has been subject to strong reputational challenges on the basis of methodological weaknesses.[97] Nevertheless, its 'Freedom in the World' benchmark has acquired substantial weight owing to its alignment with the interests and agendas of third parties such as the US government and various international organisations, which greatly magnifies its audience and influence.

In such cases, the scientific expertise or institutional status of the benchmarker may be less consequential than what other parties do with their benchmark once it is produced. This underscores the need for more nuanced analyses of how global benchmarking links up with other transnational practices, as well as how benchmarks can potentially lead to unintended consequences. To gain a richer understanding of political effects of global benchmarking, it is therefore necessary to take into account: (1) the status and history of the specific organisation or individual that has produced a given benchmark; (2) the internal mechanics of how a given benchmark is produced; (3) the distinctive characteristics and political and economic profile of the specific issue being benchmarked; and (4) the authority and credibility that third party users can invest in benchmarks when they align with other political interests and agendas.

Conclusion

Global benchmarks are inspired by frequently overlapping normative values and agendas, which are then translated into ratings, rankings, and measurements for a given category of conduct, institutional design, or outcome. They are designed to promote distinctive forms of transnational behaviour and transnational organisation by enabling symbolic judgements of performance that are expressed through numerical values. These numerical values create information shortcuts that facilitate non-expert comparisons of 'good' or 'bad' performance by radically simplifying both context and complexity. Thanks in part to the popular and political appeal of numbers, benchmarks have emerged as a key practice for both promoting and codifying many different agendas and interests, and for either legitimating or challenging a diverse range of global actors and transnational activities. While it is in the interests of benchmarkers to rhetorically appeal to models of neutral, methodical, and technocratic assessment, their activities and outputs will always be inherently political.

Global benchmarking raises a number of critical questions for IR scholars and future IR research. Benchmarking is now routinely deployed as a tool of governance

---

[96] Sabino Cassese and Lorenzo Casini, 'Public regulation of global indicators', in Kevin E. Davis, Angelina Fisher, Benedict Kingsbury, and Sally Engle Merry (eds), *Governance by Indicators: Global Power through Quantification* (Oxford: Oxford University Press, 2012), pp. 468–9.

[97] Homolar, 'Benchmarking human security'; see also Christopher G. Bradley, 'International organizations and the production of indicators: the case of Freedom House', in Sally Engle Merry, Kevin E. Davis, and Benedict Kingsbury (eds), *The Quiet Power of Indicators: Measuring Governance, Corruption, and Rule of Law* (Cambridge: Cambridge University Press, 2015), pp. 27–74.

and knowledge production across a wide range of transnational policy arenas, and there are important differences between the four main types of benchmarking that we have identified and the political impact they can have in world politics. Moreover, the growing use of global benchmarks as a tool for constructing (at least the appearance of) authoritative expertise, and for extending public and private authority over distant entities, has increased the need to connect theories of how power operates indirectly in the international realm to explanations of how and why such efforts are – *or are not* – successful at achieving their intended ends.

Accordingly, we suggest that a new research agenda for the study of global benchmarking should take on board the following lines of inquiry: How are benchmarking practices defended and legitimated, and among which audiences and in the context of which markets for activism and advocacy? Why do specific benchmarks gain traction, both among target actors and third parties, while others fail to secure an audience? Why and how does a specific benchmark have an impact in one country while remaining inconsequential in others that share broadly similar features? What types of activities and effects do global benchmarks tend to obscure or conceal? How can we better understand the long-term consequences and costs associated with benchmarking in relation to contested issues such as responsibility, accountability, and private governance? How does the practice of global benchmarking revitalise or deepen existing IR literatures relating to transnational advocacy networks, global governance and governmentality, transnational actors, rational design and cooperation, and the politics of expertise? The various contributions to this Special Issue should not only help scholars to better understand the politics of numbers and normative agendas in global benchmarking, they should also help us to in turn ask better questions about how and where the practice of global benchmarking fits within broader patterns, processes, and theories of International Relations.

Human security benchmarks: Governing human wellbeing at a distance

ALEXANDRA HOMOLAR*

Abstract. When the United Nations Development Programme formally introduced the concept of human security in 1994, it was widely celebrated as a long-overdue humanist alternative to orthodox models of security. Today, human security is a buzzword for describing the complex challenges that individuals and communities face in achieving safety and wellbeing in an insecure world. This article directs attention away from the emancipatory and empowering qualities commonly ascribed to human security to explore, instead, the specific role of benchmarking within the wider human security agenda. The main focus here is on the ways in which human life has been operationalised, measured, and classified to create indicators that permit judgements about individual security and insecurity. The article argues that although a single global human security benchmark has yet to be established, the main indices used as performance metrics of human insecurity have produced a narrow understanding of what it means to live a 'secure' life and have reinforced the state as the main focal point of international security governance.

Dr Alexandra Homolar is Associate Professor of International Security in the Department of Politics and International Studies at the University of Warwick and is Principal Investigator on the 'Enemy Addiction' Research Project, funded by a three-year ESRC Future Research Leaders Fellowship. Prior to joining Warwick, Alexandra held research positions at Yale University, Johns Hopkins University, and the Peace Research Institute Frankfurt. Her work has been published in journals such as the *European Journal of International Relations*, *Journal of Strategic Studies*, *Review of International Political Economy*, and *Critical Discourse Studies*.

So we rely on statistics in order to build and maintain our own model of the world.
<div align="right">Dudley Seers (1983)[1]</div>

For as long as we are unable to put our arguments into figures, the voice of our science … will never be heard by practical men.
<div align="right">J. A. Schumpeter (1933), p. 12.[2]</div>

Introduction

In 1994, the United Nations Development Programme (UNDP) Human Development Report[3] (HDR) made a major contribution to reshaping the international security

---

[1] Dudley Sears, *The Political Economy of Nationalism* (New York: Oxford University Press, 1983), p. 130.
[2] Joseph Schumpeter, 'The common sense of econometrics', *Econometrica*, 1:1 (1933), pp. 5–12 (p. 12).
[3] United Nations Development Programme (UNDP), *Human Development Report (HDR) 1994* (New York and Oxford: Oxford University Press, 1994), available at: {http://hdr.undp.org/sites/default/files/reports/255/hdr_1994_en_complete_nostats.pdf} accessed 5 August 2014.

agenda of the 'New World Order'. The document offered a clear vision of how the people-centric concept of 'human security' should replace the traditional focus on conflict between states, the protection of state borders, and military solutions to security problems. Human security was envisioned as an all-encompassing emancipatory concept that could help to address the many causes of human vulnerability across the globe, including violent conflict, resource deprivation, human rights violations, and environmental change. For many observers and practitioners of global politics, the notion of human security held the promise of a new normative order that could transcend the ideological straitjacket and state-centric orientation of the Cold War international system.[4]

Despite ongoing debates over the meaning, scope, and utility of the concept,[5] the idea of human security has since been widely celebrated as offering a long-overdue humanist alternative to traditional security governance with the potential to both empower and protect individuals.[6] Critics, however, have pointed to the orthodox qualities of the concept and sought to raise awareness of the counterproductive practical effects of the human security agenda.[7] The downsides identified here include the securitisation of everyday life, as well as the norms, power relations, and universalising claims that human security both contains and promotes.[8] As such scholarship has shown, the emancipatory qualities commonly ascribed to the concept of human security cannot simply be accepted at face value.[9] It is within this context that this article contributes to the debates on what human security *does*.

Specifically, the article interrogates the process of translating the abstract *notion* of human security into concrete measures of a liveable life, which lies at the centre of making judgments about the *state* of human security. Contemporary political and academic debates have already begun to shine the spotlight on how the concept could and should be operationalised to quantify its multiple dimensions of insecurity

---

[4] Roland Paris, 'Human security: Paradigm shift or hot air?', *International Security*, 26:2 (2001), pp. 87–102; Mary Martin and Taylor Owen, 'The second generation of human security: Lessons from the UN and EU experience, *International Affairs*, 86:1 (2010), pp. 211–24.

[5] See, for example, Nicholas Thomas and William T. Tow, 'The utility of human security: Sovereignty and humanitarian intervention', *Security Dialogue*, 33:2 (2002), pp. 177–92; Alexander J. Bellamy and Matt McDonald, 'The utility of human security: Which humans? What security? A reply to Thomas and Tow', *Security Dialogue*, 33:3 (2002), pp. 373–77; Edward Newman, 'Critical human security studies', *Review of International Studies*, 36:1 (2010), pp. 77–94; J. Peter Burgess, 'The ethical challenges of human security in the age of globalization', in Moufida Goucha and John Crowley (eds), *Rethinking Human Security* (Chichester: Wiley Blackwell and UNCESCO, 2008); Ralph Pettman, 'Human security as global security: Reconceptualizing strategic studies', *Cambridge Review of International Affairs*, 18:1 (2005), pp. 137–50.; Astri Suhrke, 'Human security and the interests of states', *Security Dialogue*, 30:4 (1999), pp. 265–76; P. H. Liotta and Taylor Owen, 'Why human security?', *Whitehead Journal of Diplomacy and International Relations VII: Winter/Spring* (2006), pp. 137–55.

[6] Keith Krause, 'Building the agenda of human security: Policy and practice within the human security network', *International Social Science Journal*, 59:1 (2008), pp. 65–79 (p. 78).

[7] Nik Hynek and David Chandler, 'Introduction: Emancipation and power in human security', in David Chandler and Nik Marhia (eds), *Critical Perspectives on Human Security: Rethinking Emancipation and Power in International Relations* (Abingdon: Routledge, 2011), pp. 1–10.

[8] See Ryerson Christie, 'Critical voices and human security: To endure, to engage or to critique?', *Security Dialogue*, 41:2 (2010), pp. 169–90; Tara McCormack, 'Human security and the separation of security and development', *Conflict, Security and Development*, 11:2 (2011), pp. 235–60. On human security as 'relation of governance' and as 'principle of formation' see Mark Duffield and Nicholas Waddell, 'Securing humans in a dangerous world', *International Politics*, 43:1 (2006), pp. 1–23; Kyle Grayson, 'Human security as power/knowledge: the biopolitics of a definitional debate', *Cambridge Review of International Affairs*, 21:3 (2008), pp. 383–401.

[9] See Tara McCormack, 'The limits to emancipation in the human security framework', in Chandler and Hyrek (eds), *Critical Perspectives*, pp. 99–113; see also David Roberts, 'Human security, biopoverty and the possibility for emancipation', in Chandler and Hyrek (eds), *Critical Perspectives*, pp. 69–82.

and threat.[10] This article both contributes to and goes beyond such considerations as it zooms in on the normative assumptions about the core ingredients for a 'liveable life', upon which any attempt to measure human security is implicitly based. In particular, it explores the process of normalising and essentialising human existence through the lens of global benchmarking practices, which has thus far been sidelined in the discussions over the utility and impact of the human security agenda.

As the editors of this Special Issue highlight, global benchmarks have emerged as a key tool 'for extending public and private authority over distant entities'.[11] The practice of benchmarking human security converts controversial international development and security policy agendas, complex social phenomena at the domestic level, and normative concepts about human progress into a legible and technocratic terminology that does not reflect their intense contestation. The benchmarking process entails two main steps. In the first step, the concept of human security is disaggregated into distinct categories and indicators that enable the ordinal representation of its complex constituent parts. In the second step, universal standards are constructed and applied to human security indicators to enable the comparative analysis and assessment of human vulnerability across issue areas in numerous countries over time.

The article is structured into two main parts that reflect the process of human security benchmarking. Part one critically discusses this process of operationalising human security with a primary focus on unpacking the indicators used in the 1994 HDR. While the Report is commonly associated with the launch of the human security framework, there has been less appreciation of how it also established and promoted a global agenda for how we map and measure people's living conditions and through what categories. Part two of the article shows that while thinking about human security has generally moved beyond the 1994 HDR, the indicators contained in the Report have continued to inform the practice of benchmarking of human security. Given the current lack of a global 'gold standard' in human security performance metrics and index rankings, the focus here is on a set of seemingly disparate indexes, which tend to be discussed in isolation: the Fragile State Index, Freedom in the World, and the Human Development Index. Although they may not intuitively reflect human security priorities, collectively these indices have provided global governance actors with established datasets, categories, and benchmarks to both judge the state of human security in the world and reinforce the standards constructed to achieve human security. The conclusion underscores the key finding that how human life has been operationalised, measured, and classified in the practice of benchmarking human security is at variance with the emancipatory political rhetoric commonly associated with human security discourses and policies. The main argument of the article is that human security benchmarking has been heavily implicated in the normalisation of controversial policy goals and the promotion of a one-size-fits-all approach to securing humans, which reinforces the state's responsibility for security, rather than challenging or supplanting it.[12]

---

[10] See Kyle Grayson, 'Human security as power/knowledge', p. 384; Mary Martin and Taylor Owen, 'The second generation'; Kirsti Stuvøy, 'Human security research practices: Conceptualizing security for women's crisis centres in Russia', *Security Dialogue*, 41:3 (2010), pp. 279–99.

[11] André Broome and Joel Quirk, 'Governing the world at a distance: the practice of global benchmarking', *Review of International Studies*, 41:5 (2015), pp. 819–41.

[12] Natasha Marhia, 'Some humans are more human than others: Troubling the "human" in human security from a critical feminist perspective', *Security Dialogue*, 44:1 (2013), pp. 19–23 (p. 23).

The politics of indicators and the human security agenda

Global humanitarian governance is typically couched in universalising terms. It is presented as an 'inherently progressive project' that strives towards 'enacting and creating a world defined by the values of humanity', one that is centred on protecting and improving the lives of the vulnerable.[13] But any attempt at 'saving strangers'[14] first requires the identification of *who* is vulnerable and *to what*; it requires an ability to observe and compare individual experiences of insecurity across multiple social and institutional contexts. Much of contemporary global humanitarian governance relies on a diverse collection of indicators to identify areas of greatest need. These indicators – and the indices and benchmarks that are based on them – serve as proxy measures for the state of the human condition.[15]

Indicators are always constructed and selected based upon *a priori* conceptions of which observable properties correspond with the phenomenon that is to be captured in a simplified and measurable way.[16] They never exist in a sociopolitical vacuum, as neutral attributes that stand ready to represent specific social phenomena as data and measurements. Human security is no exception. As the following discussion shows, the process of translating the concept of human security into a series of tangible, measurable objects ('reification') relies on operationalising normative assumptions about what constitutes 'liveable' human existence into observable and measurable categories. The indicators chosen to make judgements about the state of human security thereby function to both concretise and reproduce abstract ideas about what constitutes a 'secure' human life.

Delineating human (in)security in the 1994 Human Development Report

The release of policy reports by international actors occasionally has significant effects on the evolution of global policy agendas. Prominent examples include the World Bank's 1993 report *The East Asian Miracle* and the 2006 Stern Review commissioned for the UK government on *The Economics of Climate Change*. Few publications have had an impact as far-reaching and enduring as the 1994 Human Development Report *New Dimensions of Human Security*. The document gained its landmark status by officially introducing the notion of human security to the global policy community. While the concept of human security had been articulated earlier at the North-South Roundtable 'Economics of Peace' in Costa Rica in January 1990 and also was included in the 1992 UN Agenda for Peace,[17] the

---

[13] Michael N. Barnett, 'Humanitarian governance', *Annual Review of Political Science*, 16:1 (2013), pp. 379–98 (pp. 381, 383). See also Didier Fassin, 'Humanitarianism: a nongovernmental government', in Michel Feher (ed.), *Nongovernmental Politics* (New York: Zone Books, 2007), pp. 149–60 (p. 151).

[14] Nicholas J. Wheeler, *Saving Strangers: Humanitarian Intervention in International Society* (Oxford: Oxford University Press, 2002).

[15] See J. David Singer, 'Variables, indicators, and data: the measurement problem in macropolitical research', *Social Science History*, 6:2 (1982), pp. 181–217 (p. 186); also UNDP, *HDR 1994*, p. 3.

[16] Russel Lawrence Barsh, 'Measuring human rights: Problems of methodology and purpose', *Human Rights Quarterly*, 15:1 (1993), pp. 87–121 (pp. 91, 95–6); Ray Pawson, *A Measure for Measures; A Manifesto for Empirical Sociology* (London: Routledge, 1989), p. 40. As Singer ('Variables', p. 186) put it, 'we think and theorize in terms of concepts, and our data serve as convenient surrogates that represent them and can be subjected to statistical analysis'.

[17] 'Economics of Peace: A Summary of the North South Roundtable Session in San Jose', Costa Rica, 4–5 January 1990, available at: {http://ns-rt.org/reports/ECONOMICS%20OF%20PEACE.pdf} accessed 12 September 2014; Boutros Boutros-Ghali, 'An Agenda for Peace: Report of the Secretary-General' (UN Security Council, 1992), available at: {www.unrol.org/files/A_47_277.pdf} accessed 25 August 2014

terminology did not gain traction in the wider public discourse until after the launch of the 1994 HDR.[18]

The pivotal document disaggregated human security into seven interlinked composite parts: economic security, food security, health security, environmental security, personal security, community security, and political security. Intimately connected to the Universal Declaration of Human Rights and radically departing from economic interpretations of the human condition, the 1994 HDR put forward a multidimensional conceptualisation of human security aimed at providing a holistic approach to help pinpoint the greatest areas and sites of individual vulnerability. However, a key prerequisite to mapping the state of human security is the availability of a clear set of guidelines on what precisely the seven categories of human security contain, and what the major signposts are for 'insecurity' in each. Given its aim to address the 'growing challenge' of human security through a new, people-centric paradigm that could shift the focus of security from the state to the 'legitimate concerns of ordinary people' and 'daily lives', we might expect the Report to engage in a 'pro-people' operationalisation of human security.[19] But this was not the case.

The Report set out to highlight that since the end of the Second World War, humanity had progressed 'on several critical fronts' relevant to human security and affirmed a belief in the possibility to engineer further positive change.[20] It drew specific attention to the areas of liberty, self-determination, development, healthcare, and economic wellbeing. These were presented as universal values, despite their longstanding liberal heritage.[21] In order to illustrate the notion of 'human advance', the Report relied on established global comparative measures that conceptually privileged the state rather than individuals and their specific living conditions. For example, the evidence offered in support of human advance included the increase in the number of free countries following the end of the Colonial era, the steady rise in national wealth levels, and a decline in military expenditures after the mid-1980s, as well as an increase in the number of pluralistic and democratic regimes.[22]

This emphasis upon the state as a key reference point for understanding human progress diminished the extent to which it was possible to move beyond a statist bias from the outset. Moreover, the reliance on the state as the primary unit of analysis continued throughout the diagnosis of human vulnerability across the seven categories of human security. A notable example is 'economic security', which featured prominently in the 1994 HDR. While remaining elusive when it comes to a precise definition of economic security, the Report specified that 'an assured basic income' should form the core requirement to be achieved – rather than simply growth in economic output.[23] Human vulnerability in this area was equated with income insecurity, which was further disaggregated into the notions of insecure working conditions, underemployment, the decreasing value of nominal wages, and the lack of

---

(para. 16); see also the UNDP, *Human Development Report (HDR) 1993* (Oxford and New York: Oxford University Press, 1993), p. iii, available at: {http://hdr.undp.org/sites/default/files/reports/222/hdr_1993_en_complete_nostats.pdf} accessed 12 September 2014.

[18] This is based on a Factiva search of major English language print media for the phrase 'human security' from January 1990 to September 2014.

[19] UNDP, *HDR 1994*, pp. 4, 22.

[20] Ibid., p. 1.

[21] On development as norm, see David Williams, *International Development and Global Politics: History, Theory, and Practice* (Abington: Routledge, 2012).

[22] UNDP, *HDR 1994*, pp.1–2.

[23] Ibid., p. 24.

social security as important markers.[24] Despite extensive discussion of the economic security category, there was little attempt to operationalise a multifaceted understanding of income insecurity into measurable phenomena. 'Data limitations' were cited as the main obstacle to the conversion of the different constituent elements of income insecurity into a set of measurable economic security indicators applicable to all countries.[25] In order to make quantified inferences about individual economic wellbeing, the Report instead reverted back to the use of readily available national aggregate data as a proxy for economic security. This included a particular emphasis upon Gross National Product (GNP) per capita for developing countries and unemployment rates for industrial countries.[26]

Even if these two measures could accurately represent comparable indicators of aggregate economic conditions across different societies, they are nonetheless highly problematic for making judgements about the economic dimension of human security. This is because they draw upon datasets developed for a different purpose: they assess national economic *performance* rather than individual economic *security.* GNP per capita represents the average annual earnings per person as the average total value of all goods and services produced by a country in one year divided by the size of its population. The unemployment rate, meanwhile, shows the percentage of unemployed workers in a country as share of its total labour force.[27] The two key indicators utilised to provide an assessment of the state of 'economic' human security therefore lack an intuitive relationship to the core requirement of an assured basic income. Moreover, it remains unclear how national economic performance metrics correspond with a people-centric understanding of individual economic security, even for the datasets used. Another fundamental problem, with far-reaching political implications, is the ambiguity of the scale of measurement; that is, the failure to clearly identify the threshold between economic security and economic insecurity.[28]

Each of the components of the human security framework presented in the 1994 HDR suffered from a similar vagueness in relation to conceptualisation and operationalisation. In addition, many of the indicators selected were based on unexplained and value-laden assumptions about what factors are important in making individuals 'secure' across the different core categories of human security – which essential ingredients add up to a 'liveable life'. This was particularly problematic in the Report's human development priority of health security.[29]

Health security was defined as the ability of individuals, communities, and societies to avoid premature death. Key sources of health insecurity in developing countries were identified as common infectious and parasitic diseases, such as those linked to poor nutrition and an unsafe environment. For industrialised countries, in

---

[24] Ibid., pp. 25–6.

[25] Ibid., p. 26.

[26] Ibid. Since 1993, this has become the Gross National Income (GNI). Calculated in national currency, and then converted into US$, the GNI is 'the sum of value added by all resident producers plus any product taxes (less subsidies) not included in the valuation of output plus net receipts of primary income (compensation of employees and property income) from abroad'. World Bank, 'GNI per capita, Atlas method (current US$)', *World Bank Open Data* (2013), available at: {http://data.worldbank.org/indicator/NY.GNP.PCAP.CD} accessed 5 August 2014.

[27] The total labour force is defined as the actual number of people available for work, both unemployed and employed.

[28] The Report only notes that 0.1 % of donor countries GNP 'channelled to the poorest nations … for basic human development priorities' could help bring 'all poor nations up to at least a minimum threshold of human development'. UNDP, *HDR 1994*, pp. 4–5.

[29] Ibid., pp. 4, 6.

turn, the main causes noted were cardiovascular diseases and cancer, which were respectively linked to lifestyle choices and environmental contamination. The fundamental problem here is that this is not even a comparison between apples and oranges, but rather between apples and almonds. What is more, as was the case in the category of economic security, the unit of analysis for the main indicator of health insecurity is not specifically concerned with individual wellbeing, but instead focuses upon aggregate outcomes. The degree of health security is defined in terms of premature death at the national level by measuring the annual number of deaths attributed to specific yet different causes in developed and developing countries. With respect to attributes with a negative impact on health security at the individual level, the Report further singled out women and the poor as particularly vulnerable groups within society, specifically depicting maternal mortality rates as an important marker of health security and making an explicit, if unspecified, causal link between poverty and ill-health.[30]

At the heart of the health security section of the Report was the importance of how well societies are able to both counter the sources of insecurity identified and deal with their effects. This reflected an assumption that how a society is organised is causally related to the degree of health security in that society. A higher degree of health security was linked to a liberal understanding of social welfare provision. To measure and compare the relationship between the provision of welfare and health security, the Report listed a range of indicators, which are tied to a set of positions that tend to be associated with a progressive humanitarian agenda. These indicators included the ratio of doctors to people within and across countries, national per capita health spending levels, the percentage of people without health insurance, access to safe water, and malnutrition levels, as well as access to family planning and home care during pregnancy and birth.[31] This implied an obligation on the state to provide an environment that fosters equal opportunity for individuals to achieve health security by protecting them from disease as much as from unhealthy lifestyles, a concept that is far from being universally accepted. In addition to a paternalistic tendency to shape and govern peoples' life choices in a direction compatible with the way in which health security is defined, the Report failed to bridge the gap between this universalised conception of an entitlement to good health and the particular and differentiated needs associated with health problems between and within societies.

Another major problem with the 1994 HDR is that it relied upon tautological reasoning to provide measurements for specific indicators in one of the core categories of human security by utilising a second. Food security, for example, is one of the least-defined components of human security in the Report. Eliding the difference between the general availability of food in a country and the ways in which food is spatially distributed within a given territory, food security is understood as the entitlement of people to 'physical and economic access to basic food'.[32] This implicitly combines two distinct factors: (1) an unequal distribution of food and (2) a lack of purchasing power. The Report used notions of 'undernourishment' and 'born underweight' as observable indicators to quantify individuals' vulnerability in terms of inadequate access to food, without establishing how these indicators should be measured and

---

[30] Ibid., p. 28.
[31] Ibid.
[32] Ibid., p. 27. The 1994 HDR presents a table on 'Indicators of food security in selected countries', which are based on national aggregate datasets without discussing how they relate to and help to measure 'access to food'.

against what standard. Instead, it established a link between 'access to food' on the one hand, and 'access to assets, employment, and income security' on the other.[33] The degree of food security is thus at least partially measured by the level of economic security.

This overview of how the human security agenda was delineated in the 1994 Human Development Report illustrates the conceptual flaws and methodological problems inherent in how the various categories of human security were disaggregated and measured. A great deal of ambiguity remains over why these seven core components are essential to human security, what the relationship is between them, and how they overlap in terms of their conceptual scope and relevant indicators. And while the very idea of human security implies that a spectrum of security exists based on standards that would allow for cross-national comparison – a scale that ranges from secure to insecure – no attempt was made in the Report to clarify the high/low values at either end of this spectrum, or the various points that may mark qualitative differences in levels of human security along it.

Human security and international stability

In the twenty years since its inception, human security has developed into a potent buzzword for describing the complex challenges that individuals and communities face in achieving safety and wellbeing in an insecure world. Many international political actors, in particular from the developed world, have incorporated human security as part of their mandate and policy agenda. This includes states such as Norway, Japan, Canada, and EU member states; international organisations such as the UN and the World Bank; and non-governmental organisations, particularly aid agencies and human rights organisations.[34]

Today, human security is both a standard term in the development policy lexicon and a reference point for a broad understanding of security. It has also has emerged as a constitutive element of global humanitarian governance. An important factor in the continuing popularity of the human security concept and associated terminology is a legacy of how the 1994 HDR secured its meaning, which connected a concern with the security of individual human beings to questions of international peace and stability. This discursive link has two key facets.

First, the capability-centric problematisation of individuals' vulnerability to a wide range of threats beyond those to their immediate physical safety fused negative rights (civil and political) and positive rights (economic and social) within a single meta-concept. Articulated as universally shared concerns irrespective of their particular normative heritage, the core ingredients to a secure, liveable life for all people are today presented as critical objectives of – and challenges to – the broader

---

[33] Ibid.
[34] Paris, 'Paradigm shift'; Gary King and Christopher J. L. Murray, 'Rethinking human security', *Political Science Quarterly*, 116:4 (2001), pp. 585–610 (p. 589); Michael Barnett and Thomas G. Weiss, *Humanitarianism Contested: Where Angels Fear to Tread* (London: Routledge, 2011), pp. 26–9; cf. OECD, *Conflict, Peace and Development Co-Operation on the Threshold of the 21st Century* (Paris: OECD, 1997), available at: {http://pdf.usaid.gov/pdf_docs/Pcaaa817.pdf} accessed 5 August 2014; International Commission on Intervention and State Sovereignty (ICISS), 'Report: The Responsibility to Protect' (Ottawa: International Development Research Centre, 2001), available at: {http://responsibility toprotect.org/ICISS%20Report.pdf} accessed 5 August 2014); Commission on Human Security, 'Final Report: Human Security Now', available at: {http://reliefweb.int/sites/reliefweb.int/files/resources/91BAEEDBA50C6907C1256D19006A9353-chs-security-may03.pdf} accessed 5 August 2014.

international security agenda. They include universal primary education, literacy (in particular for women), access to primary health care and immunisation, freedom from discrimination, and access to food and family planning services, as well as safe drinking water, sanitation, and credit.[35] The emphasis on human security defined in terms of a 'universalism of life claims' speaks to the empowerment of people across the globe and the promotion of a wide range of positive and negative rights as basic human rights.[36]

The 1994 HDR offered only minimal insights into how the state of human (in)security can be measured, and what the threshold is between security and insecurity. Yet the indicators developed in the Report have helped to secure and reproduce a very specific meaning of 'the vital core' of human life within global humanitarian discourse and the wider international security agenda.[37] In the context of the metaphorical Global War on Terror, this liberal humanist-inspired construction of what makes a life worthy and 'liveable' beyond mere biological existence has gained rather than lost momentum.[38]

Second, 1994 HDR focused less on capturing how secure or insecure individual human beings are, and more on how well states perform according to the standard of human security. Concentrating attention on states' willingness and capacities to provide equal footing for individuals' life chances as well as to protect both human rights and the general quality of life opened the door for judgements about the condition of human security to be judgements levelled against particular states. The focus on 'good governance' across the different components of human security has both defined the core zones for political action and served to reconfigure the meaning of the state within the fabric of international security.[39] No longer is the state primarily seen as the principal referent of security, but it is also understood as the core provider and arbitrator of human security.

The shift of the human security agenda away from making inferences about levels of individual human vulnerability also helped to carve out new areas of responsibility for legitimate statehood. This marked a major steppingstone towards the emerging norm of conditional sovereignty, which is the idea that the *right* of states to be recognised as sovereign actors in the international arena without being subject to external interference is tied to their ability and willingness to protect their own population from harm.[40] The extent of this *responsibility* has remained a matter of debate – as has the point at which the international community should act on behalf of those affected.[41] Nonetheless, the radical departure from treating human insecurity as detached from recognised statehood has risen to prominence

---

[35] This is summarised in the '20:20 Compact for Human Development'. UNDP, *HDR 1994*, pp. 7–8; see also Liam Clegg, 'Benchmarking and blame games: Exploring the contestation of the Millennium Development Goals', *Review of International Studies*, 41:5 (2015), pp. 947–67.

[36] Ibid., pp. 13–14; also Des Gasper, 'Securing humanity: Situating "human security" as concept and discourse', *Journal of Human Development*, 6:2 (2005), pp. 221–45.

[37] On the 'vital core' of human life see *Sabina Alkire, A Conceptual Framework for Human Security* (Oxford: Centre for Research on Inequality, CRISE, 2003).

[38] See Marhia, 'Some humans are more human', pp. 21–3, 28; Michael Dillon 'Governing terror: the state of emergency of biopolitical emergence', *International Political Sociology*, 1:1 (2007), pp. 7–28.

[39] See also James Harrison and Sharifah Sekalala, 'Addressing the compliance gap? UN initiatives to benchmark the human rights performance of states and corporations', *Review of International Studies*, 41:5 (2015), pp. 925–45. Clegg, 'Benchmarking global development'.

[40] Alexandra Homolar, 'Rebels without a conscience: the evolution of the rogue states narrative in US security policy', *European Journal of International Relations*, 17:4 (2011), pp. 705–27.

[41] See, for example, Alexander Bellamy, 'The Responsibility to Protect – five years on', *Ethics & International Affairs*, 24:2 (2010), pp. 143–69.

in the post-9/11 era.[42] In contemporary international security and development discourses, the ability of the state to foster human capabilities and to provide a sociopolitical environment conducive to 'human security' is widely seen as a necessary function of a civilised society and increasingly sets the boundaries of political possibility for interventions by international actors.

The scholarly and policy debates over the utility and scope of the concept of human security continue to evolve.[43] Nonetheless, the human security indicators set out in the 1994 HDR, however vague, have served to reflect, substantiate, and reproduce the foundational norms of an international order that is centred on a liberal humanist notion of progress and which emphasises self-determination, representative government, and economic wellbeing. They have helped to translate abstract normative conceptions of what a liveable life entails and what the responsibilities of the state should be within this framework into concrete, even actionable, categories. Despite its various shortcomings, 1994 HDR thus played a pivotal role in the normative reconstruction of the post-Cold War international order. By shaping how we categorise and measure human security, the landmark report altered how we interpret and evaluate the success and failure of domestic and international practices.[44] As we shall see, this continues to reverberate in contemporary global benchmarking.

Benchmarking human (in)security through global indices

Global indices provide international actors with a powerful political weapon to rate and rank different countries in a systematic and comparative fashion, in order to establish how well they perform against predefined targets and to promote related policy agendas.[45] Indices are aggregations of a range of different indicators, each derived from a series of observed values that have been placed on a specific scale of measurement to enable a comparative analysis.[46] They are intended to monitor

---

[42] The universalism entails here 'respecting national sovereignty but only as long as nation-states respect the human rights of their own people'. UNDP, *HDR 1994*, p. 14. On human security and conditional sovereignty see Miguel De Larrinaga and Marc G. Doucet, 'Sovereign power and the biopolitics of human security', *Security Dialogue*, 39:5 (2008), pp. 517–37 (pp. 525–8).

[43] See, for example, the debates between the 'broad school', which advances joining together human physical safety, human dignity, and human development, and the 'narrow school', which focuses on individuals' safety from violent conflicts and has received a boost after the ICISS publication, *The Responsibility to Protect*. See Caroline Thomas, 'Global governance, development and human security: Exploring the links', *Third World Quarterly*, 22:2 (2001), pp. 159–75; Bellamy and McDonald, 'The utility of human security'; P. H. Liotta and Taylor Owen, 'Sense and symbolism: Europe takes on human security', *Parameters*, 36:3 (2006), pp. 85–102; Suhrke, 'Human security'; Thomas and Tow, 'Utility of human security'; see also the HDR 1994 Special Issue in *Contemporary Politics*, 21:1 (2015); Mary Martin and Taylor Owen (eds), *Routledge Handbook of Human Security* (London: Routledge, 2013).

[44] See also Rebecca Adler-Nissen and Vincent Pouliot, 'Power in practice: Negotiating the international intervention in Libya', *European Journal of International Relations*, 20:4 (2014), pp. 889–911.

[45] See Ole Jacob Sending and Jon Harald Sande Lie, 'The limits of global authority: How the World Bank benchmarks economies in Ethiopia and Malawi', *Review of International Studies*, 41:5 (2015), pp. 993–1010. Alexander Cooley and Jack Snyder (eds), *Ranking the World: Grading States as a Tool of Global Governance* (Cambridge: Cambridge University Press, 2015); Beth A. Simmons and Judith G. Kelley, 'Politics by number: Indicators as social pressure in International Relations', *American Journal of Political Science* (forthcoming 2015); Katja Freistein, 'Effects of indicator use: a comparison of poverty measuring instruments at the World Bank', *Journal of Comparative Policy Analysis: Research and Practice* (2015, online first); Kevin E. Davis, Benedict Kingsbury, and Sally Engle Merry, 'Indicators as technology of global governance', *Law and Society Review*, 46:1 (2012), pp. 71–104.

[46] Jack Brand, 'The politics of social indicators', *The British Journal of Sociology*, 26:1 (1975), pp. 78–90 (p. 79).

complex political, economic, and social phenomena for particular political purposes by expressing a set of disparate indicators on the basis of a common metric to permit assessments about conditions and trends.[47] As such, they permit the making of judgements and ultimately benchmark the quality of conduct of a unit of analysis, the design of institutions, and sociopolitical outcomes through standardised comparative measures that are linked to specific reference points.[48]

Despite frequent references to human security indicators[49] and to human security as a benchmark 'for an emerging new model of "security"',[50] no consensus has emerged on how to operationalise the concept or what the standard is for separating the high-achiever countries in the human security classroom from the under-performers. While mapping the patterns and trends of human insecurity across the globe has played a key role in the wider discourse on humanitarian governance in the international arena, measuring threats to human life and human vulnerability has thus far been characterised by the lack of a specific global human security index.[51] However, a variety of prominent and well-established global indices, which link to different dimensions of how human security has been delineated since the 1994 HDR, have served as tools and reference points for benchmarking human security. These are typically centred either on states' *capacities* to provide an environment that is conducive to human security, or how well the degree of development in different countries and regions fosters improvements in human *capabilities*. The following discussion of the differences between capacity-benchmarking and capability-benchmarking illustrates that, although they produce country ratings and rankings for different purposes, these indices have shaped what are today counted as the key performance metrics of human security.

Capacity benchmarking: State stability and freedom scores

In response to a spike in intra-state conflicts at the beginning of the post-Cold War era, the problem of what have been variously termed 'failed', 'failing', and 'fragile' states took centre stage on the international security agenda. The 1994 HDR is an early example of the classification of 'national breakdown' as one of the main risks to human security and, by extension, a critical threat to international stability.[52] Although both are highly ambiguous concepts, this link between the level of 'human

---

[47] On this process of commensuration see Wendy Nelson Espeland and Mitchell L. Stevens, 'Commensuration as a social process', *Annual Review of Sociology*, 24 (1998), pp. 313–43 (p. 315); On commensuration and global benchmarking practices see André Broome and Joel Quirk, 'Governing the world at a distance'.

[48] See also Theodore M. Porter, *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life* (Princeton: Princeton University Press, 1995), pp. 28, 45; On standards and benchmarking see Kathleen C. Dominique, Ammar Anees Malik, and Valerie Remoquillo-Jenni, 'International benchmarking: Politics and policy', *Science and Public Policy*, 40:4 (2013), pp. 504–13.

[49] UNESCO-ISS, 'Final Recommendations of the UNESO-ISS Expert Meeting on Peace, Human Security and Conflict Prevention in Africa', Proceedings of the UNESCO-ISS Expert Meeting held in Pretoria, 23–4 July 2001, available at: {www.issafrica.org/Pubs/Books/Unesco/FinalRcomm.html} accessed 5 August 2014.

[50] Dan Henk, 'Human security: Relevance and implications', *Parameters*, 35:2 (2005), pp. 91–106 (p. 92).

[51] Miguel De Larrinaga and Marc G. Doucet, 'Sovereign power', p. 528. See the HDI subsection below for attempts to develop such an index. The datasets and indexes the Human Security Report Project focuses relate to a narrow understanding of human security, see, for example, Human Security Audit (2005), available at: {www.hsrgroup.org/docs/Publications/HSR2005/2005HumanSecurityReport-Part2-HumanSecurityAudit.pdf} accessed 7 May 2015.

[52] UNDP, *HDR 1994*, p. 38.

security' in a country and the risk of 'state failure' gained further momentum in the aftermath of the terrorist attacks in the US on 11 September 2001.[53]

State failure, broadly defined as the lack of 'stateness', is regarded as a destructive force for human security because it is equated with the absence of the conditions established in the wider human security discourse as essential ingredients to normal statehood and a secure life: reliable rules and physical infrastructure for economic development and social interaction.[54] Some observers go as far as to argue that to gradually achieve human security in such societies, external 'statebuilding' interventions to create a 'legitimate, professional, and representative state … is the only way to address the problems of the modern, interconnected world'.[55] In contemporary debates on whether and how to intervene in a sovereign country's domestic affairs, the link between individual human wellbeing and the nature of social organisation at the state level is made explicit through the idea that legitimate (sovereign) statehood entails both the capacity and responsibility to protect and to provide for the domestic population. Should a state fail to do so, it is argued, the international community has a right and a duty to act. In extreme cases, this duty can include the authorisation of the use of force to ostensibly come to the rescue of the domestic population as external 'liberators'.[56]

Despite the politically contested nature of the concept of state failure, the degree to which a state is judged to be 'fragile' or at risk of 'collapse' is generally determined through an evaluation of the 'conditions that threaten the physical integrity, welfare, self-determination, and opportunities' of individual human beings.[57] Here, the most prominent assessment of state capacity is the *Fragile State Index* (FSI) (formerly named the Failed State Index) developed by the Fund for Peace and published annually since 2005. The target audience of the FSI is the international policymaking community, with the explicit aim of providing policymakers with a reliable tool to anticipate and assess problems of 'stateness', and to therefore serve as a 'first step in devising strategies for strengthening weak and failing states'.[58]

The Fund for Peace's global index quantifies the vulnerability of states to collapse or conflict, and ranks and classifies their performance on the basis of

[53] See Charles T. Call, 'The fallacy of the failed state', *Third World Quarterly*, 29:8 (2008), pp. 1491–507; Homolar, 'Rebels without a conscience', pp. 719–20; UN, 'World Summit Outcome' (2005), paras 138–9, available at: {www.unrol.org/files/2005%20World%20Summit%20Outcome.pdf} accessed 5 August 2014; United Nations Security Council (UNSC), 'Resolution 1674' (28 April 2006), available at: {www.securitycouncilreport.org/atf/cf/%7B65BFCF9B-6D27-4E9C8CD3CF6E4FF96FF9%7D/Civilians %20SRES1674.pdf} accessed 25 August 2014; UNSC, 'Resolution 1894' (11 November 2009), available at: {www.securitycouncilreport.org/atf/cf/%7B65BFCF9B-6D27-4E9C-8CD3-CF6E4FF96FF9%7D/POC %20SRES1894.pdf} accessed 25 August 2014.

[54] Marina Ottoway and Stefan Mair, 'States at risk and failed states: Putting security first', *Policy Outlook* (Carnegie Endowment of International Peace, September 2004), p. 6, available at: {www.swp-berlin.org/fileadmin/contents/arbeitspapiere/statesatrisk_ks.pdf} accessed 5 August 2014.

[55] Nate Haken and Patricia Taft, *The Dark Side of State Building: South Sudan* (Fund for Peace, 2013), available at: {http://library.fundforpeace.org/fsi13} accessed 1 August 2014.

[56] G. John Ikenberry and Anne-Marie Slaughter, 'Forging a World of Liberty Under Law, U.S. National Security in the 21st Century', Princeton Project Papers (Final Report 2006), p. 26; Alexandra Homolar-Riechmann, 'The moral purpose of US power: Neoconservatism in the age of Obama', *Contemporary Politics*, 15:2 (2009), pp. 179–96 (pp. 190–1); also US White House, *The National Security Strategy of the United States* (Washington, DC, 2002), p. 1; Ban Ki-moon, 'UN Secretary-General Report on "Responsibility to Protect: Timely and Decisive Response"' (2012), available at: {http://responsibilityto protect.org/index.php/component/content/article/176-the-un-and-rtop/4315-un-secretary-general-releases-report-on-responsibility-to-protect-timely-and-decisive-response} accessed 2 June 2014; ICISS, *The Responsibility to Protect*.

[57] Ottoway and Mair, 'States at risk', p. 2.

[58] Ibid.

these measures.[59] The FSI bases the rank order of individual states on the total scores of 12 primary indicators across 3 dimensions – social, economic, and political – which are each broken down into an average of 14 further sub-indicators. Each of the primary indicators is an aggregate measure, which rates a particular aspect of state performance on a scale from 0 for the lowest intensity (most stable) to 10 for the highest intensity (least stable). The final aggregate score achieved by a state in the overall composite index is the sum of all 12 primary indicators. State performance is then benchmarked via a scale of possible scores from 0 to 120. This is in turn divided into four zones of vulnerability: 'sustainable' (0 to 29.9), 'monitoring' (between 30 and 59.9), 'warning' (between 60 and 89.9), and 'alert' (between 90 and 120). The complex and quantitative nature of the FSI aims to signal scientific rigour and to convey an aura of objectivity of the rankings and ratings produced. The focus is nonetheless on judging state viability at the design level of policies and institutions against liberal norms of good governance and legitimate statehood. This includes evaluating the level of democracy, the quality of healthcare provision, the degree of political participation, and the characteristics of electoral processes, as well as the presence of political freedoms and civil liberties.[60]

The adherence to liberal norms of legitimate statehood helps to explain the utility of the Fragile State Index within a wider discourse on humanitarian governance. Moreover, many of the primary and secondary indicators used to construct the index map onto the components of human security outlined in the 1994 Human Development Report and its list of pointers for the risk of national breakdown.[61] For example, the primary indicator 'poverty and economic decline' is used to assess the level of economic security in a country through the lens of the ability of the state to provide for its population, and includes measurements such as GDP growth, GDP per capita, government deficits, and unemployment rates.[62] The underlying assumption here is that poverty and economic decline automatically have a negative impact on the ability of the state to provide for its citizens and increases tensions between the wealthy and the poor. The design of the FSI also makes it possible to assess the capacity of states to serve the needs of their populations in the categories of food security and health security. The 'mounting demographic pressures' indicator as a key marker of state vulnerability in the social dimension includes measurements related to diseases, mortality, and water scarcity, as well as malnutrition and food scarcity. Measurements related to water and sanitation, healthcare, and infrastructure, in turn, are an integral part of the 'progressive deterioration of public services' indicator in the political dimension of the FSI.[63] In its explanation for the recent renaming of the FSI from 'Failed State Index' to 'Fragile State Index', the Fund for Peace has made the link to the human security agenda explicit, suggesting that the goal 'has always been to help improve human security in countries all over the world'. The index, the Fund argues, is created precisely to help set policy priorities and to foster the responsibility of governments in improving livelihoods.[64]

---

[59] The discussion of the FSI indicators is based on information provided by the Fund for Peace 2014 available at: {http://ffp.statesindex.org/} accessed 25 August 2014.
[60] Fund for Peace, 'Indicators', available at: {http://ffp.statesindex.org/indicators} accessed 5 August 2014.
[61] UNDP, *HDR 1994*, pp. 32–40.
[62] Fund for Peace, 'Indicators'.
[63] Ibid.
[64] Fund for Peace, 'Name Change', available at: {http://library.fundforpeace.org/fsi14-namechange} accessed 25 August 2014.

The most widely recognised global comparison of whether the institutional designs of states empower or repress their constituents with respect to political rights and civil liberties is much older than the concept of human security itself. *Freedom in the World* (FIW) is a comparative assessment in the form of a state-centric composite index that has been produced by the US-based think tank Freedom House since 1972. It is derived from data generated through the organisation's Freedom in the World Survey (previously named the Comparative Survey of Freedom). The aim of this annual exercise is to measure and rank the degree of freedom – or democracy for that matter – as experienced by individuals in different countries across the globe. In contrast to legal guarantees of liberty enacted by governments, these 'experiences' are understood by the FIW in terms of 'real-world' rights and freedoms, or *de facto* rather than simply *de jure* rights.[65] While many of the FIW components overlap with the FSI indicators, the former centre more explicitly on questions of political rights and civil liberties, which are integral to the political component of human security agenda.

Freedom in the World applies normative standards derived from the Universal Declaration of Human Rights to rank all countries (as well as some disputed and dependent territories).[66] This is based on their performance across two complex composite indicators, political rights and civil liberties, in order to establish the degree of political freedom and civil liberties that individuals enjoy in each country (or not). These liberties are defined as universal standards that apply 'irrespective of geographical location, ethnic or religious composition, or level of economic development'. They include, for example, measuring the degree of individual rights, freedom of expression, and the rule of law for the civil liberties indicator, as well as political participation, organisational rights, and the functioning of the government for the political rights indicator. In-house and external analysts from a variety of expert community backgrounds undertake the FIW evaluation and assess countries' performance on the basis of a set of pre-defined questions for each of the secondary indicators. The sources on which this assessment is based vary significantly, and include news media reports, academic studies, and publications from civil society watchdogs and other non-governmental organisations as well as from individual professional contacts.

FIW uses a three-tiered assessment system of assigning scores, ratings, and a status for each of the two primary indicators. These in turn consist of a total of twenty-five secondary indicators – ten for the political rights dimension across three subcategories and fifteen for civil liberties dimension across four subcategories – as well as two discretionary political rights questions. A country receives a score for each of the twenty-five secondary indicators by individual analysts awarding points on a scale from 0 (smallest degree of freedom) to 4 (greatest degree). On the basis of the total scores that a country or territory receives in the first step, it is then assigned two ratings – one for political rights and one for civil liberties on a scale from 7 (smallest degree) to 1 (greatest degree). In the final step, the unweighted average is calculated from the political rights and civil liberties ratings, on the basis of which countries are awarded one status on a nominal scale from 1 to 7 that consists of three categories: free (1.0 to 2.5); partly free (3.0 to 5.0); and not free (5.5 to 7.0). This is the overall 'Freedom Rating' produced within the remits of the FIW, and it is widely used by

---

[65] Freedom House, 'Methodology Fact Sheet', available at: {www.freedomhouse.org/report/methodology-fact-sheet#.U_3uNkiPL-m} accessed 25 August 2014. See also {www.freedomhouse.org/report/freedom-world-2014/methodology#.VBVmT2Nvxjd} accessed 1 September 1014.
[66] This discussion of methodology is based on ibid.

policymakers and researchers to make judgements about the degree of freedom and democracy in countries across the globe.

Despite its longstanding popularity, FIW has a large number of well-known methodological problems.[67] A major issue is the lack in transparency of the coding process, which affects the reliability of the ratings produced and their replicability. It remains unclear, for instance, which kind of answers to the questions on the FIW checklist are translated by the analysts into what measures on the individual scales, and what the underlying theoretical basis is for this relationship between responses and measures. In addition, the source material is not clearly identified, contains subjective observations, and mixes assessments from different issue areas without the specification of their relevance to the FIW ratings. The disaggregated datasets that underlie the ratings produced in the FIW survey are also not easily accessible to researchers outside Freedom House.[68] Besides a normative bias that privileges systems of social organisation based on modern Western understandings of justice, liberty, freedom, and self-determination, this raises questions about the subjectivity of the scores assigned and the lack of rigour in the data compilation process.

Many aspects of the capacity-benchmarking done by the *Fragile State Index* and *Freedom in the World* play an important role in reinforcing the legitimacy of status quo international standards of conduct, institutional design, and policy implementation for states to achieve human security. This includes the principles of good governance, the rule of law, bureaucratic competence, and optimal socioeconomic infrastructure. The FSI and FIW have received sustained scholarly criticism that centres on questioning their methodology, including the principles of replicability and comparability, as well as the operationalisation of the primary and secondary indicators.[69] While the FIW has begun to address some of these issues and data has become more accessible and transparent, the FSI methodology remains opaque. The status of both indices as 'reputable' global benchmarks across a wide range of policy audiences nonetheless largely persists. The credibility of the country rankings produced by the FSI and FIW with third-party users is primarily rooted in their ability to provide international actors with an information shortcut to judge how well states measure up against the benchmark of stability and freedom, and as such feature prominently in contemporary discourses of human security.

Capability benchmarking: Human development goalposts

While the rankings produced in capacity benchmarking are frequently linked to key components of the human security agenda, the main global index used to measure

---

[67] See detailed Gerardo L. Munck and Jay Verkuilen, 'Conceptualizing and measuring democracy: Evaluating alternative indices', University of Illinois at Urbana-Champaign, *Comparative Political Studies*, 35:1 (2002), pp. 5–34; Diego Giannone, 'Political and ideological aspects in the measurement of democracy: the Freedom House case', *Democratization*, 17:1 (2010), pp. 68–97.

[68] Recent subcategory scores and aggregates scores have now been made available at: {www.freedomhouse.org/report/freedom-world-aggregate-and-subcategory-scores#.VBVmKWNvxjd} accessed 14 September 2014.

[69] See Raymond Duncan Gastil, 'The comparative survey of freedom: Experiences and suggestions', in Alex Inkles (ed.), *On Measuring Democracy: Its Consequences and Concomitants* (New Brunswick and London: Transaction, 1991); Munck and Verkuilen, 'Conceptualizing and measuring democracy'; Axel Hadenius and Jan Teorell, 'Assessing alternative indices of democracy', Committee on Concepts and Methods Working Paper Series, *Political Concepts*, 6 (2005), available at: {www.concepts-methods.org/Files/WorkingPaper/PC%206%20Hadenius%20Teorell.pdf} accessed 5 August 2014; Lars Carlsen and Rainer Bruggemann, 'The "Failed State Index" offers Morethan just a simple ranking', *Social Indicators Research*, 115 (2014), pp. 525–30.

human security through a focus on individual capabilities (what humans can do and be) laid the groundwork for the development of the concept of human security itself. The *Human Development Index* (HDI), created by the economists behind the landmark 1994 HDR, Mahbub ul Haq and the Amartya Sen, was presented as a new way of measuring development beyond a focus on state-centric aggregate measures of economic growth and economic means. Introduced in the first Human Development Report in 1990 – and further refined methodologically in the 1991 HDR and conceptually in the landmark 1994 Report[70] – the HDI has since become the primary comparative quantitative assessment of the state of human development worldwide. Its utility for global humanitarian governance lies in the HDI's aim to enable judgements on a multidimensional understanding of human security through the lens of the width of peoples' choices and the level of their achieved wellbeing.[71]

The *Human Development Index* seeks to assess progress in human development across three dimensions that are conceptually underpinned by a capability-based approach to poverty, which emphasises the importance of individual agency for development and wellbeing.[72] These dimensions include: (1) a long and healthy life; (2) access to knowledge; and (3) a decent standard of living. Each of these three dimensions of human development is organised into an index based on a wider set of indicators, such as life expectancy, education enrolment, and GNI per capita. The three areas of health, education, and income were identified in the 1990 Human Development Report as both the 'most critical' measurable development achievements and essential elements of human life – with political freedoms and guaranteed human rights seen as additional rather than core routes towards improving people's choices.[73]

For each dimension, the HDI develops a scale of extreme values by defining a maximum (which was previously the highest observed value in a time series and is now a value capped at a particular point) and a minimum (a subjectively fixed subsistence value, which is seen the basic requirements for a society to survive over time).[74] Each country is then assigned a value between 0 and 1 in relation to the 'goalposts' marking the high and low end of the human development scale. The value for each sub-index is calculated by abstracting the minimum value from the actually observed value of a measurement. The outcome is then divided by the result of abstracting the minimum value from the maximum value. The HDI value for a country is the geometric mean of the constituent indexes created for the three dimensions centred on different quality-of-life attributes, equally weighted. Following this process of quantification, which methodologically privileges values that are subjectively set over those that are observed, countries are ranked and categorised on

---

[70] UNDP, *Human Development Report (HDR) 1990* (Oxford and London: Oxford University Press 1990), available at: {http://hdr.undp.org/sites/default/files/reports/219/hdr_1990_en_complete_nostats.pdf} accessed 5 August 2014); UNDP, *HDR 1994*, ch. 5.

[71] Ibid., p. 10.

[72] See Amartya K. Sen, 'Well-being, agency and freedom: the Dewey Lecture 1984', *Journal of Philosophy*, 82 (1985), p. 203; UNDP, *Human Development Report 2010* (New York: Palgrave Macmillan, 2010), p. 16; also Amartya K. Sen, *Inequality Reexamined* (Cambridge, MA: Harvard University Press, 1992); Sabina Alkire, 'The capability approach as a development paradigm', in Enrica Chiappero-Martinetti (ed.), *Debating Global Society: Reach and Limits of the Capability Approach* (Milan: Feltrinelli Foundation, 2009).

[73] UNDP, *HDR 1990*, pp. 10, 12.

[74] See UNDP, 'Human Development Reports: Frequently Asked Questions', available at: {http://hdr.undp.org/en/faq-page} and {http://hdr.undp.org/en/faq-page/human-development-index-hdi#t292n52} accessed 25 August 2014.

an ordinal scale of human development – very high, high, medium, and low.[75] Like the FSI and FIW, the HDI has received strong criticism. In addition to problems with its methodological soundness, the quality of its underlying data, and problems with data updating and formula revisions,[76] this has included questioning the utility of the indicators selected to measure the state of the human condition and the basis for the setting of both indicator goalposts and human development thresholds.[77]

Over the past two decades, several attempts have been made to overcome some of the methodological and conceptual problems raised in connection with the HDI. In terms of relevance to the broad human security agenda, these include notably the Multidimensional Poverty Index, created jointly by the Oxford Poverty and Human Development Initiative and the UNDP in 2010 for the Human Development Reports;[78] the Generalised Human Development Index, offered by Satya R. Chakravarty;[79] as well as proposals to develop performance metrics specifically for human security, such as Gary King and Christopher Murray's threshold of generalised poverty,[80] and David Hastings' prototype Human Security Index (HSI).[81] However, these alternatives have not overcome the problem of conceptual operationalisation that has plagued the HDI. In all of these cases, indicators have been selected and operationalised on the basis of their apparent conceptual intuitiveness and readily available data, often in the form of complex aggregated datasets.

In the absence of a widely-accepted alternative, the HDI has remained the key global measure used to evaluate the living standards of a country's population in the wider human security discourse through the lens of how well individual states perform and what they have achieved with respect to the select goalposts identified in the individual indicators. The focus in the HDI on the pursuit of development – understood as leading a long and healthy life, acquiring formal educational qualifications, and gaining a decent income – implies a move away from a sensitivity to spatial and cultural differences in conceiving what it is that makes a life valuable, which was a marker of the capabilities-based approach to global humanitarian governance from which the HDI emerged.[82]

---

[75] The thresholds set in the 2013 HDR are now relative, with country classifications based on HDI quartiles. See detailed: UNDP, 'Technical Notes' (2013), available at: {http://hdr.undp.org/sites/default/files/hdr_2013_en_technotes.pdf} accessed 5 August 2014.

[76] Hendrik Wolff, Howard Chong, and Maximillian Auffhammer, 'Classification, detection and consequences of data error: Evidence from the human development index', *The Economic Journal*, 121:553 (2011), pp. 843–70.

[77] For a broad overview, see Milorad Kovacevic, 'Review of HDI Critiques and Potential Improvements', Human Development Research Paper No. 33 (2010), available at: {http://hdr.undp.org/sites/default/files/hdrp_ 2010_33.pdf} accessed 1 September 2014.

[78] Sabina Alkire and James Foster, 'Counting and multidimensional poverty measurement', *Journal of Public Economics*, 95:7–8 (2011), pp. 476–87; Sabina Alkire and James Foster, 'Understandings and misunderstandings of multidimensional poverty measurement', *Journal of Economic Inequality*, 9:2 (2011), pp. 289–314. The 2013 MPI is available at {www.ophi.org.uk/global-multidimensional-poverty-index-mpi-2013/} accessed 25 August 2014.

[79] Satya R. Chakravarty, 'A generalized human development index', *Review of Development Economics*, 7:1 (2003), pp. 99–114.

[80] King and Murray, 'Rethinking human security', pp. 590, 606; cf. Taylor Owen, 'Human security – conflict, critique and consensus: Colloquium remarks and a proposal for a threshold-based definition', *Security Dialogue*, 35:3 (2004), pp. 345–87.

[81] David Hastings, 'The human security index: Potential roles for the environmental and Earth observation communities', *Earthzine* (May 2011) available at: {www.earthzine.org/2011/05/04/the-human-security-index-potential-roles-for-the-environmental-and-earth-observation-communities/} accessed 25 August 2014; for the HSI see {www.humansecurityindex.org/} accessed 25 August 2014.

[82] See Martha C. Nussbaum, 'Introduction', in Martha C. Nussbaum and Jonathan Glover (eds), *Women, Culture, and Development: A Study of Human Capabilities* (Oxford: Oxford University Press, 1993), pp. 1–36 (p. 5); Martha C. Nussbaum, 'Capabilities as fundamental entitlements: Sen and social justice', *Feminist Economics*, 9:2–3 (2003), pp. 33–59; S. Charusheela, 'Social analysis and the capabilities

While there were initial disagreements between the HDI architects over the reduction of complex human capabilities to only three main indicators, these were set aside to create a comparative measure of human development with GDP-like traction.[83]

Although the HDI provides measures intimately related to core components of the human security agenda, the UNDP has begun to distance itself from the concept. Since the early 2000s, the framing of policy priorities in the UNDP's Human Development Reports, which contain the HDI and discuss related policy priorities, has devoted less attention to human security components. Instead it has reflected the UN's tendency over the past decade to subsume these within 'threat clusters' to enable the adaptation of the concept and related development goals to the post 9/11 security environment.[84] With the twentieth anniversary of the concept of human security in 2014, the UNDP acknowledged that the HDI cannot measure human security, and recognised that the concept of human security has lost its some of its initial utility in capturing human vulnerability through the way it has been used in scholarly and political discourses on global humanitarian governance.[85]

While the focus on a holistic and capabilities-oriented approach to improving human lives and achieving human progress has remained at the centre of the UNDP's agenda, it is now dominated by the language of *shocks* and *threats* to human development, including economic risks, inequality, health risks, environmental and natural disasters, food insecurity, and physical insecurity.[86] Although these categories map onto the human security concept introduced two decades ago and reaffirm the focus on states' capacities to 'empower and protect people',[87] the UNDP has initiated a major rhetorical shift toward the dichotomy of resilience/vulnerability.[88] This echoes a broader move to recast human security in terms of 'resilience' in (inter) national security discourse.[89]

The multiplier effects of global indices

Claims about the condition of human security tend to be based on state-centric measures of economic growth, development, good governance, and legitimate statehood. As we have seen, this pattern applies regardless of whether these measures are based on capacity-benchmarking or capability-benchmarking. While human security is sometimes seen as a component of human wellbeing and sometimes

approach: a limit to Martha Nussbaum's universalist ethics', *Cambridge Journal of Economics*, 33:6 (2009), pp. 1135–52.

[83] See Sakiko Fukuda-Parr, 'The human development paradigm: Operationalizing Sen's ideas on capabilities', *Feminist Economics*, 9:2–3 (2003), pp. 301–17; Sakiko Fukuda-Parr, 'Rescuing the human development concept from the HDI – reflections on a new agenda', in Sakiko Fukuda-Parr and A. K. Shiva Kumar (eds), *Readings in Human Development* (Delhi: Oxford University Press, 2004), pp. 117–24.

[84] United Nations (UN), 'A More Secure World: Our Shared Responsibility' (2004), available at: {www.un.org/en/peacebuilding/pdf/historical/hlp_more_secure_world.pdf} accessed 5 August 2014; see also UN, 'World Summit Outcome' (2005), available at: {www.unrol.org/files/2005%20World%20Summit%20Outcome.pdf} accessed 5 August 2014.

[85] UNDP, 'Human Development Report' (HDR) (2014), p. 18, available at: {http://hdr.undp.org/sites/default/files/hdr14-report-en-1.pdf} accessed 1 September 2014.

[86] UNDP, 'Frequently Asked Questions'.

[87] UNDP, *HDR 2014*, p. 5.

[88] UNDP, *Towards Human Resilience: Sustaining MDG Progress in an Age of Economic Uncertainty* (New York: UNDP, 2011).

[89] David Chandler, 'Resilience and human security: the post-interventionist paradigm', *Security Dialogue*, 43:3 (2012), pp. 213–29.

as the meta-concept,[90] in conjunction the global indices used to assess the degree of human security enable international actors to make quantified connections between vulnerability, underdevelopment, the lack of state capacity, and security. Various practices of benchmarking human security have served to correlate a lack of human development with threats to international peace and stability, in the form of the security-development nexus, the poverty-security nexus, the link between state capacity and security, and the identification of violent conflict as a barrier to development, and fostered the translation of such narratives of causation into policy agendas.[91] Yet despite the influence of the human security agenda on the evolution of the security discourse over the past two decades, the responsibility for addressing problems of 'insecurity' has remained firmly anchored in the state.[92]

The global indices used in capacity-benchmarking and capability-benchmarking to permit quantifiable inferences and comparative assessments about the state of human vulnerability serve both as key transnational advocacy tools for the monitoring agents that produce them and as data sources for policy activism and political science research. Indeed, the ability to measure can be seen as the symbolic capital of the field of global humanitarian governance. The attractiveness of reducing the complex web of challenges to human security into numbers lies in delivering easy-to-digest data chunks that can be acted upon, as well as in the perception of indices as 'efficient, consistent, transparent, scientific, and impartial'.[93] Through processes of measurement, counting, and calculation, performance metrics of human security that are derived from prominent global indices are perceived to carry a high degree of objectivity, and this can lend authority to actors who may have very little of their own.[94]

Benchmarking human security through global indices does not just conceal the ambiguous nature of the concepts and theories that these rankings are based on. It also multiplies the problems associated with making complex phenomena observable through their conversion into indicators, in particular the oversimplification of the phenomenon being measured and brushing over the norms and values that the indicator itself contains.[95] In this regard, the goal of international comparability is prioritised above contextual validity and accuracy. As 'composite indices', these global ratings are based on complex datasets that include numerous other indicators – often mixing different levels of measurement, as well as applying competing methodologies and data sources – which are aggregated into more encompassing indices. What is measured in global indices, how this is done, and which measures are conceptualised as indicators may appear as objective and apolitical analysis, but remains highly value-laden and subjective.[96]

---

[90] See King and Murray, 'Rethinking human security', p. 606.
[91] See Owen, 'Human security'. For a critique of these links see Nicholas Waddel, 'Ties that bind: DfID and the emerging security and development agenda', *Conflict, Security & Development*, 6:4 (2006), pp. 531–55; Tara McCormack, 'Human security and the separation of security and development', *Conflict, Security and Development*, 11:2 (2011), pp. 235–60; David Chandler, 'The security-development nexus and the rise of "anti-foreign policy"', *Journal of International Relations and Development*, 10:4 (2007), pp. 362–86; Maria Stern and Joakim Öjendal, 'Mapping the security-development nexus: Conflict, complexity, cacophony, convergence?', *Security Dialogue*, 41:1 (2010), pp. 5–30.
[92] De Larrinaga and Doucet, 'Sovereign power', p. 531.
[93] Porter, *Trust in Numbers*, p. 21.
[94] Ibid., pp. 8, 23, 74; Michael Ward, *Quantifying the World: UN Ideas and Statistics* (Bloomington: Indiana University Press, 2004), p. 5.
[95] Davis, Kingsbury, and Engle Merry, 'Indicators', particularly pp. 74–7.
[96] Ibid.; Ward, *Quantifying the World*, pp. 11, 25.

The practice of benchmarking human security reinforces dominant understandings of what responsible states do, what they look like, and the criteria on which they should be judged. In this respect, benchmarking human security indirectly serves as a global standard-setting instrument. It not only permits international political actors to make symbolic judgements about state performance across a wide range of issue areas that are intrinsically value-laden but have all the appearance of being value-neutral, but it also furnishes a persuasive rationale for international intervention to foster acceptance of and, ultimately, adherence to those standards. Benchmarking human security thus facilitates international policy interventions to reduce human vulnerability, which are based on a moral evaluation of states' performance, while masquerading as an objective and value-neutral practice.

Conclusion: Governing humans at a distance

This article has shown that the concept of human security as a whole and its core dimensions are hard to unpack into clear-cut indicators. The establishment of a threshold that separates 'secure' human beings from those who are 'insecure' in the different categories of human security has remained equally elusive. What human security *is* has largely been defined by pointing to situations of widely recognised human vulnerability and suffering. We understand the idea of human security only in the abstract, as the binary opposite of an egregious *lack* of security. This enables external policy interventions in countries' domestic affairs to be justified with reference to the human security agenda and the norms that are integral to it without a clear articulation of what the particular goals of such actions should be, or how they can be achieved through such interventions.[97]

Contemporary political and academic debates on human security have thus far failed to closely examine the concept in light of global benchmarking practices and the normative assumptions about 'human-ness' upon which they rest. Such an examination helps to reveal that the 'human' at the centre of global humanitarian governance is not simply a biological fact.[98] Indeed, a significant part of the history of the notion of humanity – and humanitarianism as its practical extension – is marked by attempts to conceptualise the essence of desirable human attributes. While both the core of what it means to be 'human' and what the markers are for a liveable life have always remained subject to change over time, the social construction of the civilised 'human' has typically centred around high levels of education, culture, and status across different societies, which has often been linked to processes of inclusion and exclusion.[99]

The human security agenda does not escape this political dynamic of inclusion and exclusion. The global indicators and benchmarks chosen to make judgements about the state of human security reflect and reproduce a series of normative assumptions about what constitutes a 'secure' human life, while excluding alternative attributes that lie outside the limits of this normative framework. Expressed as measurable

---

[97] Christopher J. Finlay, 'How to do things with the word "terrorist"', *Review of International Studies*, 35:4 (2009), pp. 751–74.

[98] See Michael N. Barnett, 'Humanitarian governance', *Annual Review of Political Science*, 16:1 (2013), pp. 379–98 (p. 385); Kyle Grayson, 'The biopolitics of human security', *Cambridge Review of International Affairs*, 21:3 (2008), pp. 383–401.

[99] Detailed in Costas Douzinas, 'The many faces of humanitarianism', *Parrhesia*, 2 (2007), pp. 1–28 (pp. 1–5); on the polarization of 'humanity', see Ikechi Mgbeoji, 'The civilized self and the barbaric other: Imperial delusions and the challenges of human security', *Third World Quarterly*, 27:5 (2006), pp. 855–69.

indicators, the essentialism embedded within the practice of benchmarking human security and the wider human security discourse has not only involved developing a model for what it means to be secure (or not), but also setting standards for a 'good' human life. This has helped to establish the criteria by which what is found to be normal in a statistical sense also becomes normal in a moral sense.[100] Within this context, the practice of benchmarking human security functions as a symbolic judgement on the quality of countries' conduct, capacities, and institutional design. Because those states that are judged to be underperformers become the focal point of international policy agendas, benchmarking human security functions as an indirect strategy of governance to remake the state at a distance.

Much of the political power of human security indicators and benchmarks rests on their dual role in global governance discourses. On the one hand, they serve to underwrite legitimation claims by international political actors in pursuit of specific policy agendas to change state behaviour. On the other, they act as an 'authorising force' in the creation of international legal standards, such as the emerging norm of the 'responsibility to protect'.[101] The goal of improving human life is thereby reduced to a narrow range of categories and indicators that are defined at the outset as core dimensions of human security. While the human security label has been invested with political power precisely because its meaning and scope remain contested, the power of *benchmarking* human security is the capacity to shape, normalise, and naturalise a particular understanding of what makes human life 'secure', while at the same time obscuring the power relations and political normativity that this conception of a secure human life contains. It is the power to identify, in the deceptively neutral language of technocratic assessment, the *where*, *what*, and *how* of international policy interventions in the name of securing humans.

---

[100] Detailed Ian Hacking, *The Taming of Chance* (Cambridge: Cambridge University Press, 1990).
[101] On 'authorizing force', see Habermas's discussion of Hanna Arendt's conception of political power. Jürgen Habermas, *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy*, trans. William Rheg (Cambridge, MA: MIT Press, 1996), p. 148.

Global benchmarking networks: the cases of disaster risk reduction and supply chains

TONY PORTER

Abstract. While the importance and key features of transnational benchmarking have received increased attention in recent years, the organisational contexts of this benchmarking have not yet been analysed systematically enough. Drawing on actor-network theory, this article identifies two key aspects of these contexts. The first is embeddedness, which refers to the degree to which there are networks that carry information and action from the benchmark to its contexts and back that are sustained institutionally, including by technical artefacts, objects, and routine local practices. Embeddedness is critical to the effectiveness of benchmarking. The second is publicness, which refers to the degree to which these networks exhibit expanding flows of information across widening circles of engagement, or instead are marked by control, domination, and exclusion. The article argues that benchmarking as a form of governance has an inherent impetus toward greater publicness, although this can be prevented by self-interested actors or problems associated with embeddedness. The article then explores the relevance of these conceptual points by examining cases of transnational benchmarking: disaster risk reduction and supply chain management.

Tony Porter is Professor of Political Science, McMaster University, Hamilton, Canada. His most recent books are *Transnational Financial Associations and the Governance of Global Finance: Assembling Power and Wealth* (2013), coauthored with Heather McKeen-Edwards, and his edited *Financial Regulation after the Global Financial Crisis* (2014), both with Routledge. His research has focused on institutional changes in transnational governance, with particular emphasis on global finance, private governance, and technical systems.

Benchmarking has proliferated transnationally as an instrument of governance. We are beginning to get a clearer understanding of its significance, its properties, and its variants. For instance, André Broome and Joel Quirk identify four distinctive components of benchmarking: simplification and extrapolation; commensuration (the development of common metrics); reification (the translation of complexity into stable quantifiable categories that are presumed to be universal); and symbolic judgment (which introduce assessments of relative performance or value into the process). They also identify four types of global benchmarking, each associated with different monitoring agents (national governments; international and regional organisations; profit-based institutions; and civil society actors).[1]

This article aims to add to our understanding of global benchmarking by examining the relationship between benchmarking and the varying properties of the organisational contexts within which it operates. Benchmarking is often treated as a mechanism for ranking that is constructed by certain actors to alter the conduct of other actors. For instance, the four distinctive components of benchmarking identified

---

[1] André Broome and Joel Quirk, 'Governing the world at a distance: the practice of global benchmarking', *Review of International Studies*, 41:5 (2015), pp. 819–41. See also the literature they reference.

by Broome and Quirk primarily refer to the components of the measuring device and the four types refer to variation in the actors which produce and manage benchmarks.[2] The connections these have with the contexts within which they operate are noted at times, for instance when the reification aspect of the benchmarking process leads to the neglect of cultural or historical context. They also point to the importance of third party users. However these are mostly treated as exogenous to the more central operations of the benchmarking process, and these and many other elements of the organisational contexts of benchmarking have not yet been analysed systematically enough.

Why should we care about the *contexts* of global benchmarking? I argue that understanding global benchmarking's contexts is crucial to two issues that are most central to its significance. The first issue is whether it is effective. If a researcher sponsored by a powerful actor skillfully designs a benchmark, paying careful attention to constitutive elements such as the metrics, but then leaves these sitting in a notebook on the desk, this benchmark will be ineffective. Obtaining responses from targeted actors requires linkages between the benchmarking mechanism and the targeted actors. However unlike other types of governance, the targeted actors often cannot simply be commanded to comply with benchmarks. The benchmarks may need to activate complex networks of information, resources, and action to achieve their effects, often over long distances and numerous organisational boundaries. These networks are sustained or not sustained by the contexts of benchmarking. Where the networks are sustained by a sufficiently dense context we can say that they are *embedded*.[3] The importance of these networks and contexts is often obscured by an overemphasis on the power of the actors that create benchmarks, the weakness of targeted actors, or the formal properties and content of the benchmarking mechanism, such as the ranking.

The second crucial and related issue is its publicness. We should not only care about whether a governance instrument's design allows it to achieve the outcomes associated with it, but also whether these are outcomes which contribute usefully to *good* global governance and are in the global public interest, solving global public policy problems, or instead are tools used by self-interested actors to enhance their own power or wealth. These self-interested actors can be public, such as when powerful states use benchmarking to coerce weaker ones, or private, such as when firms use benchmarking to gain competitive advantage. An example of the former is the creation in 2002 of the US Millennium Challenge Account, which set out 16 eligibility criteria that the world's poorest countries were required to meet before being considered for US aid.[4] An example of the latter is Xerox Corporation, which pioneered business benchmarking in the 1970s in order to improve its performance

---

[2] Broome and Quirk, 'Governing the world'.

[3] On embeddedness generally see Mark Granovetter, 'Economic action and social structure: the problem of embeddedness', *American Journal of Sociology*, 91:3 (1985), pp. 481–510. On its relevance to transnational networks see Tony Porter, 'Compromises of embedded knowledge: Standards, codes and technical authority in global governance', in Steven Bernstein and Louis W. Pauly (eds), *Global Liberalism and Political Order: Toward a New Grand Compromise?* (Albany: SUNY Press, 2007), pp. 109–31; Timothy J. Sinclair, 'Reinventing authority: Embedded knowledge networks and the new global finance', *Environment and Planning C: Government and Policy*, 18:4 (2000), pp. 487–502. Despite the reference to 'embedded networks' in the title of Jennifer Bair's 'Analysing economic organization: Embedded networks and global chains compared', *Economy and Society*, 37:3 (2008), pp. 339–64, the article and its summary of the literature do not discuss the role of objects or technical artefacts in this embedding, unlike the present article.

[4] Susanne Soederberg, 'American empire and "excluded states": the Millennium Challenge Account and the shift to pre-emptive development', *Third World Quarterly*, 25:2 (2004), pp. 279–302.

relative to competitors.[5] I argue below that to evaluate the publicness of a benchmark it is necessary to analyse the degree to which the flows of information and action between the benchmark and its contexts are open and collaborative, with opportunities for targeted actors to shape benchmarks, or instead are marked by control, domination, and exclusion.

By identifying the importance of embeddness and publicness in benchmarking contexts it is possible to understand better the potential and limitations of benchmarking as an instrument in global governance. This article argues that there is an inherent logic promoting an expansion of publicness in benchmarking, but that this is continually limited by the difficulties associated with embeddedness. This challenges those who see benchmarks as neutral measuring devices that are helping observe target actors' performance in a detached manner. A typical example of such ostensible detachment is the cross-national benchmarking of airports by Air Transport Research Society, which noted that 'based on a rigorous analysis of airport cost and revenue data, the ATRS benchmarking study assessed the productivity and efficiency of nearly 200 airports and selected the most efficient airports in each of the regions, as well as the most efficient airport, overall'.[6] This article challenges those who minimise the agency of those striving to meet benchmarks and see benchmarking as primarily creating and obscuring relations of power. Tore Fougner, for instance, sees country competitiveness benchmarking by the World Economic Forum and the International Institute for Management Development as helping produce neoliberal states. Susanne Soederberg makes a similar argument about the California Public Employees' Retirement System's Permissible Country Index, which provides benchmarks for socially responsible investment.[7] It will be argued that a public impetus in benchmarking is evident in the history of business benchmarking, which displays ever-widening circles of engagement over time, but also in the two quite different cases of contemporary transnational benchmarking that are examined in the second part of this article.

The article draws upon actor-network theory (ANT) in analysing the organisational contexts of benchmarking. A distinctive feature of actor-network theory is the attention it devotes to the material aspects of extended chains of action. This emphasis on the characteristics of networks is *not* intended to suggest that interests are not important, but rather to bring out the types of connections that are needed for interactions to occur, whether these are associated with the self-interested exercise of coercive power, the positive-sum interests that rational choice theorists have extensively analysed, or quite different motivations, such as empathy.

To help identify the role of organisational contexts in benchmarking I have chosen two very different cases of transnational benchmarking to examine. The first is the benchmarking of the management of the uncertainties and risks associated with

---

[5] Christopher Bogan and Dan Callahan, 'Benchmarking in rapid time', *Industrial Management* (2001), pp. 28–33.

[6] Air Transport Research Society, 'Press Release' (18 July 2014), available at: {http://www.atrsworld.org/docs/ATRS%202014%20Airport%20Benchmark%20study%20-%20Press%20release.pdf} accessed 29 August 2014.

[7] Tore Fougner, 'Neoliberal governance of states: the role of competitiveness indexing and country benchmarking', *Millennium: Journal of International Studies*, 37:2 (2008), pp. 303–26 and Susanne Soederberg, 'Socially responsible investment and the development agenda: Peering behind the progressive veil of non-financial benchmarking', *Third World Quarterly*, 28:7 (2007), pp. 1219–37. These types of effects are also discussed by Broome and Quirk, 'Governing the world'.

global disasters, such as tsunamis. The second is the benchmarking of performance in global supply chains. Both are analysed by examining documents associated with them, drawing on commentary about them in professional, business and scholarly publications. The cases have been chosen to maximise the variation in the types of benchmarking involved. We shall see that this variation is most crucially associated with variation in contexts rather than the properties of the benchmarking mechanisms or the power of the types of actors associated with them.

Conceptualising the organisational contexts of transnational benchmarking

Benchmarking and its contexts

Broome and Quirk have defined benchmarking as 'an umbrella term for a wide range of comparative evaluation techniques – such as audits, rankings, indicators, indexes, baselines, or targets – which systematically assess the performance of actors, populations, or institutions on the basis of standardised measurements, metrics, and rankings.'[8] This emphasises the production of the benchmark itself. This leaves a number of other processes associated with benchmarking outside its implied boundaries in this definition. This includes the networks that carry the effects of the benchmark to the actors that it targets. Although one could use the term *benchmarking* to refer to the larger processes it involves, it is useful analytically to separate out the narrower properties of the benchmark itself from the larger processes that the benchmark draws on and activates. Therefore this article refers to these larger processes as the *contexts* of benchmarking, and retains the narrower conception of benchmarking consistent with Broome and Quirk's specification above.[9]

Rather than treating benchmarking as neutral measuring device or a tool used by powerful actors, this article treats benchmarking as a more active and autonomous relay in a set of networks that extend well beyond the interaction between the benchmark's architects and targets. This emphasis on networks is consistent with the centrality of governance at a distance for benchmarking's distinctiveness. One can certainly imagine a ranking system similar to benchmarking in a close hierarchical relation when a superior punishes a subordinate for not achieving a target that the former has created. However this is a very conventional form of governance, and benchmarking is only new, distinct, and meaningful as category if it involves something more: a significant degree of autonomy and discretion for those seeking to improve their performance, and governance at a distance, where the pressure to comply comes not from a superior's direct threat of punishment, but from a pressure external to the direct relationship between the benchmark's architect and target, such as market pressure, peer pressure, or the willingness of third parties to provide resources to the target actor.

[8] Broome and Quirk, 'Governing the world'.
[9] In an extensive review of the usage of the term 'benchmarking', Jeffrey Alstete finds much variation and confusion. He distinguishes between 'performance measurement' and 'true benchmarking', which includes 'a plan that is implemented and continuously studied for ongoing improvement' (p. 182). This emphasis on recursive implementation is consistent with the approach I advocate in this article, but I reserve the term 'benchmarking' here for what Alstete labels 'performance measurement'. See Jeffrey W. Alstete, 'Measurement benchmarks or "real" benchmarking?', *Benchmarking: An International Journal*, 15:2 (2008), pp. 178–86

Certain transnational benchmarking processes appear in practice to be more collaborative and horizontal than hierarchical.[10] Benchmarking in such processes is used by their architects to address their own collective problems. Benchmarking by the Organisation for Economic Cooperation and Development (OECD) can take this form when it involves the sharing of best practices among states of similar degrees of power. An example is the OECD's well known Programme for International Student Assessment (PISA), which measures student performance in elementary and secondary grades.[11] These more democratic processes are circular, recursive, and reminiscent of the well-recognised stages of policy processes in democracies, such as agenda setting, problem identification, decision, implementation and evaluation, with the outcome of policy then feeding into the next round of the process.

In the case of this more horizontal benchmarking the stages start with the identification of the problem; followed by the identification of best practices; and then the formulation of metrics and scales (the part of the process that Broome and Quirk's definition especially focuses on); the collection of performance data; the circulation of comparative performance data in a manner designed to elicit changes in performance; and finally revision of goals and measures used in this process. This then leads back recursively to the beginning stages of the process. In the more democratic and horizontal variants of this process each stage is accompanied by interaction between benchmarking and its contexts. The early stages involve channels or networks to bring information into benchmarking, and the later stages require networks to carry information out and activate pressures for compliance.

Analysing these interactions between benchmarks and the contexts through which the networks associated with them are sustained matches the historical shift in the study of domestic public policy to integrate analysis of policy networks outside the formal structures of government, and is consistent with the growing complexity and density of transnational governance mechanisms and the transnational communities they interact with.[12] This article's emphasis on benchmarking contexts is complementary to Ole Jacob Sending and Jon Harald Sande Lie's contribution to this Special Issue, which suggests that practices can precede benchmarks, and which highlights the authority relationship between benchmarks and those who are benchmarked.[13] However this article puts more emphasis on the materiality of the networks involved, rather than the more abstract qualities of relationships of authority. This recursive process is very similar to, and overlaps with, the processes that have been identified as *experimentalist governance* as developed and specified by Charles F. Sabel and Jonathan Zeitlin.[14] Their experimentalist governance model is more elaborate, defined, and complete as a form of governance than is benchmarking, which is a smaller and more mobile element of governance. Nevertheless both address the distinctive challenges of governance in fast-paced, horizontal, and complex settings that are common today, including in transnational settings.

[10] Nico Groenendijk, 'EU and OECD benchmarking and peer review compared', paper presented at Third EUCE Annual Conference, 'The EU in a Comparative Perspective', Dalhousie University, 26–8 April 2009, available at: {http://doc.utwente.nl/71955/1/Groenendijk09eu.pdf} accessed 14 September 2014.

[11] Kerstin Martens and Dennis Niemann, 'When do numbers count? The differential impact of the PISA rating and ranking on education policy in Germany and the US', *German Politics*, 22:3 (2013), pp. 314–32.

[12] Marie-Laure Djelic and Sigrid Quack (eds), *Transnational Communities: Shaping Global Economic Governance* (Cambridge: Cambridge University Press, 2012).

[13] Ole Jacob Sending and Jon Harald Sande Lie, 'The limits of global authority: How the World Bank benchmarks economies in Ethiopia and Malawi', *Review of International Studies*, 41:5 (2015), pp. 993–1010.

[14] Charles F. Sabel and Jonathan Zeitlin, 'Learning from difference: the new architecture of experimentalist governance in the EU', *European Law Journal*, 14:3 (2008), pp. 271–327 (p. 276).

Actor network theory

Actor network theory (ANT) provides a number of useful ways of understanding and highlighting the roles of these networks and how they can be embedded, which I briefly summarise here.[15]

First, as previously noted, a distinctive feature of ANT is the attention it devotes to the material aspects of extended chains of action. While norms or ideas may be important, they cannot travel from one place to another without being carried physically, by a human, or a non-human object such as a paper document. Non-humans, whether constructed by humans or present independently of humans in nature, can have effects that are similar to those associated with human actors, as suggested by the ANT concept of 'actant' which can refer to both.[16] A key concept in ANT is translation,[17] which goes beyond the conventional meaning to include the ways that the humans and non-humans who constitute a network can modify ideas and actions as they travel through that network. The materiality of the networks and contexts through which the influence and action associated with a benchmark travel is crucial to understanding whether and how that benchmark will be effective.

The associations that must be established between actants in a network if action is going to travel over distances or humans are going to think or act in coordinated, structured ways are often fragile or unpredictable, but they can also be strengthened by being embedded in objects, such as written documents. An 'immutable mobile' is an object with knowledge embedded in it, such as a checklist or a schedule, which can circulate and organise humans as it travels. Digital technologies now can allow such objects to be updated and reconfigured in real time, creating 'mutable mobiles'.[18] These features of networks are important in understanding how power can be reinforced or disrupted in them.

Bruno Latour has argued strongly against the common tendency to attribute human ideas or actions to a mysterious invisible social substance or structure, and has insisted on the importance of tracing the travels of ideas and actions in the manner

---

[15] The goal is not to provide a comprehensive overview of ANT, nor to imply that it does not have certain limitations of its own. On ANT generally, see Bruno Latour, *Reassembling the Social: An Introduction to Actor-Network Theory* (Oxford: Oxford University Press, 2005) and Barbara Czarniawska, 'Bruno Latour: an accidental organization theorist', in Paul Adler et al. (eds), *The Oxford Handbook of Sociology, Social Theory & Organization Studies: Contemporary Currents* (Oxford: Oxford University Press, 2014), pp. 87–105. For an early application of ANT to international relations see Gavin Kendall, 'Global networks, international networks, actor networks', in Wendy Larner and William Walters (eds), *Global Governmentality: Governing International Spaces* (London: Routledge, 2004), pp. 59–75. For more recent applications, see for example, Peter Vandergeest, Stefano Ponte, and Simon Bush, 'Assembling sustainable territories: Space, subjects, objects and expertise in seafood certification', *Environment and Planning A*, 47 (2015), pp. 1–19, doi:10.1177/0308518X15599297, and Jacqueline Best, *The Rise of Provisional Governance? The Politics of Failure and the Transformation of Global Development Finance* (Cambridge: Cambridge University Press, 2014).

[16] Michel Callon, 'Some elements of a sociology of translation: Domestication of the scallops and the fishermen of St. Brieuc Bay', in John Law (ed.), *Power, Action and Belief: A New Sociology of Knowledge?* (London: Routledge and Kegan Paul, 1986), pp. 196–33; John Law, 'On the methods of long-distance control: Vessels, navigation and the Portuguese route to India', in John Law (ed.), *Power, Action and Belief*, pp. 234–63.

[17] Bruno Latour, 'The powers of association', in John Law (ed.), *Power, Action and Belief*, pp. 264–80 (p. 266); Bruno Latour, *Science in Action: How to Follow Scientists and Engineers through Society* (Cambridge, MA, Harvard University Press, 1987), pp. 132–6. See also Michel Callon, 'Some elements of a sociology of translation'.

[18] On immutable mobiles see John Law, 'On the methods of long-distance control', pp. 234–63. On mutable mobiles see John Law and Annamarie Mol, 'Situating technoscience: an inquiry into spatialities', *Environment and Planning D: Society and Space*, 19:5 (2001), pp. 609–21 (p. 619).

noted above, which can then lead to the assembling of things seen as social.[19] As Latour puts it, 'it is possible to remain faithful to the original intuitions of the social sciences by redefining sociology not as the "science of the social" but as the *tracing of associations*'. This is a contrast to diffusion models, where ideas are imagined to travel from one human mind to another often without consideration of the objects, such as documents or computer networks, which are required for this type of effect to occur.[20] It also differs from some uses of the concept of governmentality, which was developed by Michel Foucault. Governmentality is a decentred form of power that involves control at a distance, and the regulation of self-regulation, where subjects are produced who align their thinking and action with the imperatives of power.[21] While some governmentality approaches bring in mundane artefacts and objects in way that is compatible with ANT, in general the emphasis is on larger political rationalities which are often assumed to be implemented unproblematically, missing the complications that ANT highlights. As Peter Gibbon and Stefano Ponte have noted, 'while rationalizations based on these models may be invoked in national or corporate economic strategies over long periods, it by no means follows that they are implemented on the ground or, where implemented, that they lead to material changes'.[22] Diffusion models and governmentality approaches are compatible with approaches that focus more narrowly on benchmarks without considering adequately the context of benchmarking. ANT is useful for highlighting the importance of these contexts.

In analysing power, ANT also looks to the capacity of actors to 'enrol' networks.[23] Power is therefore not a resource that is possessed by an actor to deploy at will. Rather it requires the alignment of existing networks towards the actor's goal. These existing networks may have their own purposes, but may come together, sometimes by chance, to constitute a larger and more powerful assemblage. This decentralised conception of power and influence is consistent with the networks and contexts that render benchmarks effective. Barbara Czarniawski has argued for the concept of 'action nets', which are emergent patterns of activity that often precede in time the emergence of actors, networks, or organisations, challenging ANT's assumption that actors and networks are preconditions for patterns of activity.[24] However, clearly actors in networks or organisations can also initiate actions. With regard to benchmarking, these points reinforce the importance of contexts for constituting successful benchmarking arrangements, while also showing that there are many opportunities for existing and emergent networks to be assembled into

---

[19] See Latour, *Reassembling the Social*.

[20] On diffusion see Beth A. Simmons and Zachary Elkins, 'The globalization of liberalization: Policy diffusion in the international political economy', *American Political Science Review*, 98:1 (2004), pp. 171–89 and David Levi-Faur, 'The global diffusion of regulatory capitalism', *Annals of the American Academy of Political and Social Science*, 598 (2005), pp. 12–32.

[21] For applications of governmentality to international relations see, for example, Larner and Walters *Global Governmentality*.

[22] Peter Gibbon and Stefano Ponte, 'Global value chains: From governance to governmentality', *Economy and Society*, 37:3 (2008), pp. 365–92. They very usefully look at documents, benchmarks, and training practices, and the ways they can fail, but they do not consider sufficiently the role of technical artefacts and non-humans more generally. The contrast between governmentality and ANT is also discussed in Tony Porter, 'Making serious measures: Numerical indices, peer review and transnational actor networks', *Journal of International Relations and Development*, 15 (2012), pp. 532–57.

[23] For an insightful ANT-inspired application of the concept of enrolment in global governance see John Braithwaite and Peter Drahos, *Global Business Regulation* (Cambridge: Cambridge University Press, 2000).

[24] Barbara Czarniawska, 'On time, space and action nets', *Organization*, 11:6 (2004), pp. 773–91.

benchmarking arrangements, without expecting the architects of benchmarks to orchestrate everything.

Despite this decentralisation, key nodes in networks can at times exercise powerful effects over the networks as a whole. The concept of an 'obligatory passage point' is a point in the network through which actions must pass.[25] Examples include a scholarly reference that all articles must cite, a computer portal through which all electronic communications must travel, or an official office, through which all documents must pass. Such an obligatory passage point cannot be peremptorily assigned by a powerful actor, but instead depends on the properties of the network. A successful benchmark can be treated as an obligatory passage point.

ANT has devoted particular attention to the way that scientific laboratories, maps, or other technologies for measuring and observing, achieve their effects. The process of successive refinements, or reifications, to reduce the complexity of the world outside to a few small and manageable relationships on a piece of paper, allows those who possess this paper to rearrange these relationships and then bring them back out into the world, thereby obtaining leverage over it.[26] The flow of information in and back out is carried by networks, and the process is very similar to that associated with effective benchmarks.

Effectiveness and publicness

How can these three ideas from ANT be used to understand transnational benchmarking better? Together they strongly suggest that we need to examine the distinctive material properties of the networks and contexts to understand how benchmarks work, and more specifically to analyse the two crucial issues discussed earlier in this article.

The first of these issues was the effectiveness of benchmarking. ANT suggests that the effectiveness of benchmarks depends on the properties of their contexts. A key factor that contributes to effectiveness is whether the context of benchmarks is sufficiently dense to support the networks needed for benchmarks to be effective. If they are, we can say that they are *embedded*. Benchmarks that travel no farther than a file on their architect's desk are not embedded, while ones that are integrated into dense distant networks with substantial material presence have a higher degree of embeddedness.[27] This embeddedness involves the type of material artefacts or objects highlighted by ANT. This goes beyond Mark Granovetter's well-known use of embeddedness since his is a more abstract conception, and much of the research he has inspired does not consider the relevance of variations in the objects that carry or mediate network ties.[28] By itself effectiveness does not imply that a benchmarking arrangement is ethically good, since it may enable powerful actors to more effectively

---

[25] Callon, 'Some elements'.
[26] Bruno Latour, *Science in Action*, p. 236.
[27] For the application of *embeddedness* to transnational networks see the references cited in fn. 3 above.
[28] Granovetter, 'Economic action and social structure', argues that actions are embedded in 'social relations'. However, since the embedding of these in objects is not discussed, the reader may assume that these are reliant on the type of mysterious social substance that Latour has criticised. Similarly the research reviewed in Mark Granovetter, 'The strength of weak ties: a network theory revisited', *Sociological Theory*, 1 (1983), pp. 201–33, does not consider technical artefacts or relationships mediated by communications technologies as significant, perhaps because in part because the internet had not yet developed. As noted in fn. 3 above, the review of the global value chain literature by Bair in 'Analysing economic organization' does not consider these either.

dominate weaker ones. However regardless of its ethical character, it is useful to know whether a benchmarking arrangement is effective or not.

The second issue, which has more ethical significance, is publicness, the degree to which benchmarks are helpful in addressing public policy problems or instead are tools used by self-interested actors to enhance their own power and wealth, perhaps at the expense of other actors or the public interest. Publicness is a useful criterion for assessing the normative qualities of a governance mechanism and the degree to which it addresses collective problems or instead is used to promote a particular interest.[29] In the past it was common to associate 'public' with governmental institutions and 'private' with business,[30] but this distinction has been undermined by the performance of public functions by private firms and by the capture of public institutions by private interests. As Jacqueline Best and Alexandra Gheciu have argued, such complications can be addressed by a practice approach which sees publics as assembled around matters that are recognised by the community involved as being of common concern, rather than as bounded spheres that have certain enduring social properties themselves.[31] This is consistent with ANT. As Latour has put it, 'neither the public, nor the common, nor the "we" exists; they must be brought into being'. This is not incompatible with the large literatures on global publics which have been inspired by but have gone beyond Jürgen Habermas's seminal work on the public sphere,[32] moving from his focus on a single national liberal public sphere to recognise multiple and often transnational publics, including counter-publics which resist dominant ones.[33]

Typical practices that are likely to promote outcomes that are in the public interest include participation, transparency, accountability, and democracy.[34] These are procedural ways of identifying the public interest that help avoid the arbitrary character of judgements by researchers on whether outcomes are in the public interest – who are researchers to make such a judgment on behalf of the public? Such a procedural approach has its own challenges. For instance, as Daniel Naurin has argued, transparency may lead to an increase in self-interested rhetoric as those involved in governance opportunistically play to an audience in a way that they would not if they could formulate policies more collaboratively in private. However Naurin also points to the recurrent presence of policy rationales formulated with reference to the public interest, with and without transparency.[35] These complications can be mitigated by not focusing too narrowly, formulaically or precisely on one measure of the public character of benchmarking processes, but instead to look for clues that help

---

[29] This article's approach to the public is consistent with its treatment as a set of social practices in Jacqueline Best and Alexandra Gheciu (eds), *The Return of the Public in Global Governance* (Cambridge: Cambridge University Press, 2014).

[30] These terms have also long had alternative meanings, such as the association of private with intimacy and the public with popular opinion or civil society.

[31] Jacqueline Best and Alexandra Gheciu, 'Theorizing the public as practices: Transformations of the public in historical context', in Best and Gheciu (eds), *The Return of the Public*, pp. 15–44.

[32] Jürgen Habermas, *Structural Transformation of the Public Sphere*, trans. Thomas Burger (Cambridge: MIT Press, 1989).

[33] See, for example, Vikki Bell, 'Introduction: the potential of an "unfolding constellation": Imagining Fraser's transnational public sphere', *Theory, Culture & Society*, 24:4 (2007), pp. 1–5, and the Special Issue that accompanies it.

[34] See Walter Mattli and Ngaire Woods, 'In whose benefit? Explaining regulatory change in world politics', in Walter Mattli and Ngaire Woods (eds), *The Politics of Global Regulation* (Princeton: Princeton University Press, 2009), pp. 1–43.

[35] Daniel Naurin, *Deliberation Behind Closed Doors: Transparency and Lobbying in the European Union* (Colchester: ECPR Press, 2007).

assess whether a benchmarking process is primarily run for the benefit of its architects and at the expense of those it targets, or if instead it is a tool used to solve collective problems in a more democratic fashion. Who participates or is listened to in formulating and revising benchmarking, how open the process is, and how responsive it is to concerns of those it targets are all useful in comparing the publicness of benchmarking in different locations or points in time. With benchmarking, it is unlikely that the presence or absence of these characteristics can be assessed by only examining the relative power of the architects or targets of benchmarks, whether these are from the public or private sector, or the benchmarks' formal properties. Instead it is necessary to explore the broader networks that carry flows of information and action between the benchmark and its contexts.

Greater publicness has intrinsic ethical benefits, helps ensure that a governance arrangement contributes to solving collective global problems rather than simply advancing particular interests, and is likely to improve effectiveness since the information flows it involves promote learning and accountability. In the theoretical literatures relevant to transnational benchmarking, Sabel and Zeitlin's experimentalist governance model[36] as discussed above, is a highly recursive and inclusive system of collective target-setting and identification of best practices. It exemplifies a type of benchmarking with high degrees of publicness. In contrast, the application of Foucault's concept of governmentality to benchmarking highlights ways that it can reinforce relations of domination, since target actors are subject to a system of power and governance that they may resist but not create, reducing the publicness of the benchmarking mechanism, as does the presence of 'capture', where powerful business actors self-interestedly manipulate a governance system that is supposed to be public.

While it is easy to see how powerful actors or practices can incorporate benchmarking as an instrument which expands the control of subordinate actors, playing one target actor off against another as they vie to meet benchmarks that the powerful have formulated, I argue that the inherent logic of benchmarking pushes it more strongly in the direction of expanded publicness. In contrast to forms of governance based on command-and-control, benchmarking is enhanced by comparisons across greater numbers of actors; by more flows of information among them; by more active engagement by them; and by a greater willingness to rework existing structures in pursuit of collective goals. These are all elements of publicness as well. Therefore a decision to engage in benchmarking and the search for and adoption of best practices associated with it tends to incorporate a governance mechanism based on a logic that encourages expanding publicness.

This impetus towards publicness is consistent with the history of private-sector benchmarking, which has evolved from evaluating products and services to include learning and implementation of best practices.[37] There has been a shift of learning from the individual and group level towards more collective learning that changes organisational cultures. This has involved an increasing integration of benchmarking and strategic planning, change management, and process reengineering.[38] Over time benchmarking has also increasingly involved looking to other companies for process

---

[36] Sabel and Zeitlin, 'Learning from difference'.

[37] Rambabu Kodali and G. Anand, 'Benchmarking the benchmarking models', *Benchmarking: An International Journal*, 15:3 (2008), pp. 257–91 (p. 266). A key work in the evolution of benchmarking is G. H. Watson, *Strategic Benchmarking: How to Rate your Company's Performance against the World's Best* (New York, John Wiley & Sons, Inc., 1993).

[38] Bogan and Callahan, 'Benchmarking in rapid time'.

lessons, with a corresponding increase in the sharing of information.[39] Collaborative benchmarking where companies work jointly on their interactions is a very recent development.[40] In other words, the range of participants, the type of information shared, and the degree of structural change involved has continually expanded over time, shifting benchmarking from a quite narrowly focused tool used by managers to a form of governance, with expanding circles of participation.

However this expansion in publicness in benchmarking is certainly far from guaranteed. The challenge of embedding the necessary networks and flows of information can inhibit this logic. As well, powerful actors and practices can work to reinforce their control of subordinate actors by incorporating the aspects of benchmarking that assist this.[41] An analysis of the contexts of any particular case of benchmarking and the degree of embeddedness and publicness that this reveals is important in assessing the degree to which that benchmarking mechanism primarily works to enhance the power of dominant actors or instead to expand publicness and address collective problems. Despite this variation in the publicness of benchmarking, I hypothesise that the way that its inherent logic encourages publicness is sufficiently strong that it will be a prominent and visible tendency in transnational benchmarking. The cases presented in the next section provide support for this hypothesis.

Disaster risk reduction and supply chain benchmarking

Disaster risk reduction benchmarking: the problem

Over the past century there has been a dramatic increase in the number of natural and technological disasters.[42] During the 1990s, 188 million people per year on average were affected by disasters, six times the comparable number for conflict. Over 90 per cent of disaster fatalities occur in the developing world.[43] A strong theme in the disaster literature is the increasing recognition over time of the social and political dimensions of all disasters.[44] A related trend at the UN and elsewhere is the shift in emphasis from relief and recovery to disaster risk reduction and the fostering of resilience.[45] 'Natural' disasters can be socially constructed, such as when new human settlement changes the landscape and triggers floods or when people move into earthquake zones. Builders may have an interest in concealing risks and governments may prefer to postpone investments in infrastructure which might mitigate disaster risk. Benchmarking is a way to address these problems.

---

[39] Paula Kyrö, 'Revising the concept and forms of benchmarking', *Benchmarking: An International Journal*, 10:3 (2003), pp. 210–25.

[40] Togar M. Simatupang, and Ramaswami Sridharan, 'A benchmarking scheme for supply chain collaboration', *Benchmarking: An International Journal*, 11:1 (2004), pp. 9–30.

[41] Larner and Le Heron detect a similar broadening of benchmarking into a form of governance, but without the publicness discussed here. See Wendy Larner and Richard Le Heron, 'Global benchmarking: Participating "at a distance" in the globalizing economy', in Larner and Walters (eds), *Global Governmentality*, pp. 212–32.

[42] See EM-DAT, 'The International Disaster Database', available at: {www.emdat.be} accessed 30 August 2014.

[43] UNISDR, *Living with Risk: A Global Review of Disaster Reduction Initiatives* (New York: United Nations, 2004), ch. 2, p. 45.

[44] For a useful overview of the literature on disasters see Uriel Rosenthal, R. Arjen Boin, and Louise K. Comfort (eds), *Managing Crises: Threats, Dilemmas, Opportunities* (Springfield: Charles C. Thomas, 2001).

[45] UNISDR, *Living with Risk*, ch. 1, p. 8.

Disaster risk reduction benchmarking: the benchmarking design, architects,
and target actors

There are many transnational benchmarking initiatives relevant to disaster risk reduction. I focus here on the most prominent one, the Hyogo Framework Self-Assessment, which is associated with the United Nations and national governments.[46] The Hyogo Framework Self-Assessment is part of a larger 'Hyogo Framework for Action (HFA) 2005–2015: Building the Resilience of Nations and Communities to Disasters' which was adopted by 168 countries at the World Conference on Disaster Reduction and endorsed by the UN General Assembly in 2005. The UN International Strategy for Disaster Reduction (UNISDR) was tasked by the conference with creating a systematic review mechanism. Five priorities for action were identified.[47] The HFA Review is a voluntary self-assessment of progress towards these priorities. An online monitoring and review tool was launched in 2007 and between by 2013, 146 countries had participated in a review, although not all reports were comprehensive. There are local, national, and regional reports, with the latter seeking to address trans-boundary disasters. The targets of the benchmarks are therefore all three levels of government, although 'individual States are mainly responsible for the implementation of disaster risk reduction and for the development and application of national indicators'.[48] National reports are organised by 'national focal points' which produce the quantitative ratings by aggregating the subjective assessments of multiple stakeholders. The local assessments measure performance relative to 'Ten Essentials' for making cities disaster resilient. For instance Essential 7 is 'Ensure education programmes and training on disaster risk reduction are in place in schools and local communities.'[49]

The HFA 'National Progress Reports' webpage provides a 'Query Builder' that permits the visitor to 'Create your own view of country national HFA progress by selecting, sorting and comparing the data and information you wish to see.'[50] For instance the Interim national progress report for 2011–13 provided by Kenya provides a score of '3' for Kenya's level of progress achieved on 'Priority for action 1' which is to 'Ensure that disaster risk reduction is a national and a local priority with a strong institutional basis for implementation.' Level 3 signifies 'Institutional commitment attained, but achievements are neither comprehensive nor substantial.' A set of responses to 'Key Questions and Means of Verification' are then provided to document the basis of this score.[51]

---

[46] Other interesting disaster risk reduction benchmarks include the Risk Mapping Index produced by DARA, a humanitarian NGO that works in more than sixty crisis countries, and the World Economic Forum's work on national disaster resilience indicators, which is primarily associated with private sector actors such as insurance companies.

[47] For instance #1 is 'Ensuring that DRR is a national and local priority, with a strong institutional basis', available at: {http://www.preventionweb.net/files/32916_implementationofthehyogoframeworkfo.pdf} accessed 28 August 2014, p. 3

[48] UNISDR, 'Indicators of Progress: Guidance on Measuring the Reduction of Disaster Risks and the Implementation of the Hyogo Framework for Action' (2008), {http://www.preventionweb.net/files/2259_IndicatorsofProgressHFA.pdf} accessed 21 August 2014.

[49] UNISDR, 'The 10 Essentials for Making Cities Resilient', available at: {http://www.unisdr.org/campaign/resilientcities/toolkit/essentials} accessed 21 August 2014.

[50] Available at: {http://www.preventionweb.net/english/hyogo/progress/reports/index.php} accessed 20 February 2014.

[51] 'Kenya: National progress report on the implementation of the Hyogo Framework for Action (2011–2013) – Interim', an HFA Monitor update published by PreventionWeb (8 January 20130, available at: {http://www.preventionweb.net/english/countries/africa/ken/} accessed 20 February 2014.

In a November 2013 report, UNISDR set out further proposals for 'a new system of indicators to measure progress in disaster risk management'. This was in response to a call for such indicators from the fourth session of the Global Platform for Disaster Risk Reduction, with 3,500 participants from 172 countries.[52] The indicator system would include measures of losses from disasters, together with resilience indicators, disaster risk management policy indicators, and underlying risk drivers indicators. The underlying risk drivers include 52 indicators in categories such as 'economic and fiscal structure', 'poverty and social vulnerability', and 'overall governance'.

The contexts of disaster risk reduction benchmarking

The above discussion has provided an overview of the architects, targets, and design of transnational HFA disaster risk reduction benchmarks. Each of these is important in understanding the significance and functioning of the benchmark. For instance the inclusion of transnational regions and local governments in the HFA benchmarks indicates that an effort is being made to link the scale and locations of the target actors to the character of the problems the benchmark aims to address. The provisions for summarising, displaying, or reporting on national progress reports are an indicator of the importance of the reactions of the target actors to this benchmarking initiative. There are also important design questions that have been discussed. For instance the UNISDR has specified the 'characteristics of good indicators', such as 'the indicator measurement should enable comparison over the different life-cycle stages of the policy or project, as well as between different policies or projects'.[53]

Despite the importance of these elements of disaster risk reduction (DRR) benchmarking it becomes quickly evident in observing them that their functioning depends crucially on the contexts of this benchmarking initiative, including the degrees of embeddedness and publicness. The success of this benchmarking depends on the strengthening of linkages between the transnational benchmarks themselves, and the most local, specific, and material aspects of disaster risk mitigation and resilience. This strengthening requires that the linkages be embedded in a robust and sustainable way that brings together humans and objects in a recursive process of feedback loops from the benchmark to local activities. It also requires that levels of participation, flows of information, and relations of accountability be strengthened, expanding the reach of the networks, bringing in an expanding variety of governmental and non-governmental actors: that the publicness of the networks be increased. In other words, the UNISDR DRR benchmarks are configured to be an active relay in circular, robust, and open networks that span the stages set out in the conceptual section above.

The importance of these circular flows is readily apparent in the UNISDR documents. For instance the data that provide the inputs to the benchmarks are not easily available but depend on a chain of information and action: 'obtaining measurements and maintaining reliable data series requires dedicated expertise and resources and can be costly. The data and methodologies upon which indicators and benchmarks depend are inevitably limited and imperfect.'[54] However the networks involve far more than this. The benchmark aims to solicit the active engagement of multiple stakeholders in a variety of ways and stages of the circle. This is evident for

---

[52] UNISDR, 'Towards the Post-2015 Framework for Disaster Risk Reduction: Indicators of Success: A New System of Indicators to Measure Progress in Disaster Risk Management' (21 November 2013).
[53] UNISDR, 'Indicators of Progress', p. 7.
[54] Ibid., p. 18.

instance in the large numbers and variety of types of actors involved in the international UN-sponsored meetings that initiated, assessed and revised the HFA. In the implementation of the benchmarking in national or local contexts the goal similarly is to mobilise large numbers and varieties of stakeholders to alter their own conduct, to assess their own progress relative to the benchmarks, to help identify appropriate measures and ways to meet benchmarks, and to hold officials accountable, pressuring them to meet benchmarks. For instance, Phase I of the Local HFA Tool has five steps for convening stakeholders for the self-assessment that is its focus. Phase II then focuses on a review and revision process with stakeholders, and Phase III calls for the report to be posted to the web.[55]

The importance of the networks that elicit local engagement is also apparent in the 2007–13 summary of the national HFA reports. This summary indicated that progress had been made in recognising the importance of DRR and related issues such as the importance of gender, and in establishing policy and legal frameworks. However, the report noted that such frameworks 'do not automatically translate into effective DRR' (p. 7). It calls for a new post-2015 framework for DRR which 'should offer incentives for scaling up DRR activities to the point where they can make a perceptible material difference in peoples' lives' (p. 8). This will require greater local outreach, as well as overcoming resource and capacity challenges.[56] In other words, the networks had begun to extend as far as national officials and legislators, but not to citizens' daily activities: more networks with more embeddedness and publicness are needed.

Online templates play an important role in increasing embeddness and publicness of networks at the national and local levels. Such electronic infrastructures embed the benchmarks and the interactions with them in computer software and hardware, establishing new links between the relevant stakeholders. They can create a degree of automaticity in the flow of information, enhancing publicness. For instance the English language version of the Local HFA Self-Assessment Tool discussed above is done through an on-line template, and the final step is 'Make the report public by clicking "submit".'

The outcomes that the benchmarking seeks to promote require the best practices that it identifies to be extended to very local and specific activities. For instance a report featured on the UNISDR website on Timor-Leste, one of the countries most exposed to extreme natural disaster risk, notes that disaster resilience there requires a mix of road improvements and changes in farming practices, since most of its people are subsistence farmers in villages cut off from other parts of the country. Changes in choices of crops and management of the soil are important in increasing the capacity for farming to recover after disaster. Changes at the political level are important, but these need to be combined with changes in the ongoing mix of human and natural activity on the land. The article on the UNISDR website creates a link in embedding networks needed for the benchmarking to connect with this type of change, but to be successful the network needs to extend in other ways to such local mixes of humans and natural actions.[57] This can include NGOs. For instance, a coalition of

---

[55] 'Local HFA Self-Assessment Tool Second Cycle', available at: {http://www.preventionweb.net/english/ hyogo/hfa-monitoring/local/?pid:73&pih:2} accessed 21 August 2014.

[56] UNISDR, 'Implementation of the Hyogo Framework for Action: Summary of Reports 2007–2013' (2013), available at: {http://www.preventionweb.net/files/32916_implementationofthehyogoframeworkfo. pdf} accessed 20 February 2014.

[57] Integrated Regional Information Networks, 'Boosting Farm-Level Disaster Resilience in Timor-Leste' (8 August 2014), UNISDR PreventionWeb, available at: {http://preventionweb.net/go/38886} accessed 22 August 2014.

development and environmental NGOs, including the Netherland Red Cross, Care, Cordaid, the Red Cross/Red Crescent Climate Centre, and Wetlands International, has been engaged with the post-2015 Hyogo Framework in order to strengthen the links between local communities and ecosystems and the global HFA process.[58] This too involves strengthening the embeddedness and publicness of networks, and the way they engage with humans, technical artefacts, and other material aspects of the connections, such as the interactions of farmers, crops, and soil. The extension of the indicators to underlying risk factors such as poverty which was noted above is a further indication of the expanding circles of engagement of the HFA process.

Transnational supply chain benchmarking: the problem

As production has become more global the management of transnational supply chains has become extremely important.[59] Brands have been damaged by supply chain problems resulting in hazardous products, environmental damage or scandalous labour conditions. Supply chain management (SCM) has become more challenging as the supply chain has become more global, more complex, involving larger numbers of firms and products, more complicated products and production processes, faster manufacturing and delivery times, and leaner operations with less room for error. As an industry participant has noted, 'Companies have systematically taken costs out of the system, but they have systematically added risks … Supply chains are much more fragile than before.'[60] In contrast to an earlier time when a single firm could impose a set of requirements on its own logistics employees or a few suppliers, there is a far greater need for collaborative standards that cut across multiple firms, functions, and industries. As well, efficiencies in supply chains are crucial for the competitive success of firms involved in them. Transnational supply chain benchmarking is designed to address these issues.

Transnational supply chain benchmarking: the benchmarking design, architects, and target actors

There are a number of benchmarking systems that have been designed for supply chain management. This case will focus in turn on the most widely used one, the

---

[58] Partners for Resilience, 'Policy Brief for post-2015 Framework for Disaster Risk Reduction: Key Messages on Building Resilient Communities' (20 May 2014), available at: {http://www.climatecentre. org/downloads/File/PFR/PfR%20Policy%20brief%20HFAII%20May%202014.pdf} accessed 28 August 2014.

[59] There is a large literature on supply chains. Some of this is primarily oriented to business, which I draw upon below. There are several approaches that provide complementary analysis of power and organisation with supply chains, with some linking of these to larger contextual analysis of the positioning of states, regions, or social classes in the global political economy. For overviews see Peter Gibbon, Jennifer Bair, and Stefano Ponte, 'Governing global value chains: an introduction', *Economy and Society*, 37:3 (2008), pp. 315–38, and the articles in the accompanying Special Issue, including Bair, 'Analysing economic organization', and Gibbon and Ponte, 'Global value chains'. See also Stefano Ponte and Timothy Sturgeon, 'Explaining governance in global value chains: a modular theory-building effort', *Review of International Political Economy*, 21:1 (2014), pp. 195–223. Larner and Le Heron, 'Global benchmarking' and Gibbon and Ponte, 'Global value chains', have analysed benchmarking in supply chains in ways that are complementary to the approach developed in the present article, but the present article addresses the significance of the contexts of benchmarking more explicitly, as well as focusing on more recent developments.

[60] Jim Lawton, vice president of Open Ratings, which provides supply chain analysis, quoted in David Biederman, 'Looking for trouble: Companies use benchmarking to help manage risks of increasingly complex supply chains', *Journal of Commerce* (2006), p. 50.

Supply-Chain Operations References (SCOR) model, and a newer one that has been developed for faster retail supply chains, the Consumer Goods Forum Global Scorecard.[61]

The SCOR model was developed by the Supply-Chain Council (SCC) and first released in 1997.[62] As one study noted, 'it is widely acknowledged as the quasi-industry standard for supply chain management'.[63] The SCC was a non-profit organisation founded by two consulting firms in 1996, initially with 69 member-companies. In 2014, it had 406 members, including 89 large global corporations such as Dupont, Ernst & Young and Microsoft, 155 smaller firms, and 162 non-profits, mostly universities and research institutes. The business members were mostly industrial, business services and transportation services rather than retail.[64] In 2014, it merged with APICS, which is the leading professional association for supply chain management (SCM) and provider of SCM training and professional certifications, with nearly 40,000 members in more than 200 countries.[65]

The SCOR model provides clear and detailed specifications of types of supply chain activities and processes. This facilitates benchmarking across processes, functions, firms, and industries. SCOR 11.0 distinguishes six standard processes: plan; source; make; deliver; return; and enable. The source-make-deliver sequence captures the flow of material in and out of each firm and can appear multiple times in a supply chain. 'Return' captures the reverse flow of defective products. 'Plan' involves the coordination of these bi-directional flows. 'Enable', which has been added more recently, is defined in the model as processes which 'support the realisation and governance of the planning and execution processes of supply chains', such as the management of rules, compliance, and contracts. It has standards for metrics in five categories: reliability; responsiveness; agility; cost; and asset management efficiency. The first three are 'customer facing' and the final two are 'internal facing'.

SCORmark is a benchmarking system based on the SCOR model that is provided by APICS SCC, in alliance with a PwC subsidiary and that complements benchmarking efforts that individual firms can do with the SCOR model. SCORmark benchmarking assesses the performance of a firm relative to selected peer groups. The firm identifies targets, collects metrics data, and enters the data through a web portal. The firm is compared to peer firms in the five categories specified above.[66]

[61] For a comparison of leading models see Douglas M. Lambert and Sebastián J. García-Dastugue, 'An evaluation of process-oriented supply chain management frameworks', *Journal of Business Logistics*, 26:1 (2005), pp. 25–51. For supply chain benchmarking of corporate social responsibility, see Genevieve LeBaron and Jane Lister, 'Benchmarking global supply chains: the power of the "ethical audit" regime', *Review of International Studies*, 41:5 (2015), pp. 905–924. The differences between this and the benchmarking that is the focus of this article are discussed below.

[62] Gordon Stewart, 'Supply-Chain Operations Reference Model (SCOR): the first cross-industry framework for integrated supply-chain management', *Logistics Information Management*, 10:2 (1997), pp. 62–7.

[63] Faskia Bete Georgise, Klaus-Dieter Thoben, and Marcus Seifert, 'Adapting the SCOR model to suit the different scenarios: a literature review and research agenda', *International Journal of Business and Management*, 7:6 (2012), pp. 2–17 (p. 2).

[64] 'Members by Membership Type', available at: {https://supply-chain.org/membership-by-type} and 'Members by Industry', available at: {https://supply-chain.org/membership-by-industry} both accessed 24 August 2014.

[65] APICS, 'APICS: Built on a Foundation of Excellence', available at: {http://www.apics.org/about/overview/mission} and 'Supply Chain Council to Merge with APICS', APICS Press Release, available at: {http://www.apics.org/news-landing-page/2014/04/30/supply-chain-council-to-merge-with-apics} both accessed 24 August 2014.

[66] This information on SCORmark is taken from materials on the SCORmark webpage, including, available at: {https://supply-chain.org/f/SCORmark_SampleReport.pdf} accessed 26 August 2014.

The second supply chain benchmarking system that I examine here is the Global Scorecard that is sponsored by GS1 and the Consumer Goods Forum, and supported operationally by IBM Global Business Services. In contrast to the SCORmark, the Global Scorecard is more focused on integrating retail-oriented supply chains with consumer preferences.

One aspect of the type of integration that the Global Scorecard seeks to promote is the barcodes and Radio-Frequency ID (RFID) technologies that are used to track products through to the point of sale to consumers.[67] RFID technology has far greater capacity to store data and communicate with computer systems than do conventional barcodes – indeed there could be a unique code assigned to each individual item produced worldwide. Barcodes, RFID, and many other related standards are managed by GS1, a non-profit association with more than 1.2 million member companies in about 148 countries. As GS1 notes, 'billions of times each day, the GS1 System of Standards connects physical things such as products, locations, logistics units, and assets with information about those things supplied by GS1 Communication Standards'.[68] An example of the integration of GS1 standards with the Scorecard is this performance measure: '% of sales with synchronized master data between trading partners via the GS1 global data synchronization network (GDSN)'.[69]

The Consumer Goods Forum (CGF) is also a sponsor of the Global Scorecard. The CGF was created in 2009 through a merger between the Global Commerce Initiative (GCI), CIES (the food and consumer goods industry association), and the Global CEO Forum.[70] The CGF brings together senior executives and managers from seventy countries representing companies with combined sales of EUR 2.5 trillion.[71]

The Global Scorecard describes its benefits as follows:

This set of tools allows companies to measure their current performance and capabilities using a globally standard set of measures. This allows the company to benchmark itself against others within the same global corporation and against peers in the same country, distribution channel or product category. Performance gaps can be identified and plans put in place to close the gaps or even break away from the crowd.[72]

The Scorecard is an 'on-line action planning tool' that assesses the performance of companies with regard to collaboration within supply chains. It is associated with the 'Efficient Consumer Response' (ECR) initiatives to 'develop a responsive, consumer-driven system in which manufacturers, brokers and distributors work together to maximize consumer value and minimize the supply chain cost'.[73] Different levels of the Scorecard are offered, with the Entry Level asking more basic questions that can

---

[67] Mohsen Attaran, 'RFID: an enabler of supply chain operations', *Supply Chain Management: An International Journal*, 12:4 (2007), pp. 249–57.

[68] GS1, 'The Value and Benefits of the GS1 System of Standards' (2011), available at: {http://www.gs1.org/docs/GS1_System_of_Standards.pdf} accessed 9 March 2012.

[69] IBM Global Business Services in collaboration with CGF, 'Make Your Supply Chain More Efficient by Using GS1 Global Standards: Findings of the 2011 Consumer Goods Forum Compliance Survey', Executive Brief (2011), p. 4, available at: {http://www.globalscorecard.net/live/download/GBW03166-USEN-00.pdf} accessed 1 March 2014, p. 7.

[70] 'The Consumer Goods Forum', available at: {http://www.traceone.com/en/partners/detail-partners/partner/the-consumer-goods-forum} accessed 24 February 2014.

[71] CGF, 'The Organisation', available at: {http://www.theconsumergoodsforum.com/about.aspx} accessed 24 February 2014.

[72] GCI and IBM, 'The GCI Global Scorecard', Brochure (2006), p. 3, available at: {http://globalscorecard.gs1.org/gsclive/download/user_guide.asp} accessed 30 August 2014.

[73] Canadian ECR, 'What is Efficient Consumer Response?', available at: {http://www.ecr.ca/en/Default.htm} accessed 1 March 2014. Information on the Global ECR initiative is available at: {http://www.ecr-all.org/} accessed 1 March 2014.

be entirely addressed on-line by collaborators, and the Intermediate Level starts on-line and then requires a face-to-face workshop. The Intermediate Level is more detailed and can integrate action items to the assessment functions. A Full Global ECR Scorecard is for the most integrated and important relationships.

In addition to collaborative partners being able to assess and improve their own specific relationship, the initiative produces an annual Global Scorecard Survey which produces global benchmarks. From its start in 2004, the annual numbers of scorecards included in the survey have increased year by year to 9,099 in the 2011 survey, representing 8,790 companies from 62 countries with total revenues of US$ 2.25 trillion.[74]

The contexts of global supply chain benchmarking

While there are many supply chain benchmarking frameworks, it is clear that each one's effectiveness depends not only on its architects, targets and design, but on the degree to which it is embedded in the operations of supply chain networks. The types of widening circles of engagement that publicness involves are also important.

The importance of embeddedness in networks is most obvious in the case of the Global Scorecard, since this requires that firms work the standards it measures into the IT and logistics systems, along with the physical objects that are transported through the supply chains. An example is the creation of an RFID tag embedded in a package by a manufacturer's supplier, which then seamlessly communicates with harmonised digitised systems as the package travels through one firm to another, towards the consumer. However the success of the SCOR model is also closely related to its ability to interact with the practical organisation of the flow of products through the supply chain, and integrating information systems with the SCOR model also provides significant benefits.[75] As the SCC's chief technology officer has noted, the SCOR model 'is a process framework that allows for rapid and reliable end-to-end descriptions of work, material, and information flows in supply chains. In a real sense, it's a supply chain dictionary.'[76] Its benefits depend on the ability of the firms using it to actually rework all of these flows to improve performance.

The ever-widening circles of participation that publicness involves are also evident in both benchmarking systems. This is not uniform. Some of the challenges to the expansion of these systems are technical, such as how to integrate the standards with IT systems. Some are associated with the search for political or commercial advantage. For instance, access to key information about the SCOR model is restricted to paid-up members. Nevertheless, there are tendencies with both systems that match aspects of publicness. The comparisons involved in both the SCORmark and the Global Scorecard are more effective with larger numbers of firms, especially where a larger set of participants is needed to have enough firms in subcategories of particular types of supply chains for comparisons across these subcategories to be meaningful. In the case of the Global Scorecard, the benefits for each firm of adopting standardised GS1 communications technologies is increased with each additional firm that adopts them as well, since it expands the potential for automating transactions

---

[74] IBM, 'Make your Supply Chain more Efficient', p. 4.
[75] See Liene Liepina, and Marite Kirikova, 'SCOR Based ISS Requirements Identification', in Witold Abramowicz, Leszek Maciaszek, and Krzysztof Węcel (eds), *Business Information Systems International Workshops and BPSC International Conference* (Poznan, Poland: Springer, 2011), pp. 232–43.
[76] Joe Francis, 'Team building with the SCOR model', *Supply Chain Management Review* (2007), pp. 60–5.

across multiple firms. Additionally, in both cases, the capacity to improve processes associated with the flow of goods through the supply chain is increased with the active engagement of everyone involved with that flow. In the case of the SCOR model, one of its big benefits is to focus everyone to work collaboratively on the standard processes that shape supply chain performance (plan; source; make; deliver; return; and enable) rather than organisational silos.[77]

The expanding circles draw in customer preferences as well, and for retail supply chains these can include public goods such as environmental sustainability. Sustainability is a major focus of the CGF-sponsored 2016 Future Supply Chain project, which states 'What does the current focus on sustainability have to do with on-shelf availability and costs in the physical supply chain for consumer goods? The answer is "everything".'[78] The integration associated with the Global Scorecard increases traceability for individual products, with the potential for increased accountability in corporate social responsibility. However, as Genevieve LeBaron and Jane Lister have shown, corporate social responsibility benchmarking, which seeks to impose an external non-commercial purpose on the internal profit-oriented operations of corporations, has been less effective than the more collaborative business activities discussed in this article.[79]

The expanding circles can also be seen in the pre-history of the Global Scorecard, which can be traced back to innovations pioneered by Walmart. Walmart built its success in becoming the world's largest corporation in part on its aggressive and early integration of point-of-sale data with data warehousing and forecasting. Walmart has long been known for the most ambitious supply chain benchmarking effort in the private sector.[80] Its relationships with its thousands of global suppliers have been managed through its electronic data interchange (EDI) system. Wal-Mart's ruthless management of its suppliers and its incessant demands to exceed previous performance measures are legendary.[81] In the past this system effectively set industry standards for supply chains, led by Walmart as the dominant firm.[82] However over time it has become more collaborative. Walmart was an early participant and leader in the Collaborative Planning, Forecasting, and Replenishment (CPFR) initiative sponsored by the Voluntary Interindustry Commerce Solutions Association (VICS) beginning in the late 1990s, which aimed for increased synchronised 'e-collaboration' on the management of barcode-related and other data.[83]

The Global Commerce Initiative, which initiated the Global Scorecard and was subsequently merged into the Consumer Goods Forum, was launched in 1999 to further develop global supply chain standards. Its most prominent working group, on EDI, was co-chaired by Kraft Foods and Walmart (Nairn 2000). The goal of the standards was to create wider, more participatory, and more open forms of

---

[77] Francis, 'Team building'.
[78] CGI and Capgemeni, '2016 Future Supply Chain' (2010), available at: {http://www.futuresupplychain.com/downloads/} accessed 27 August 2014.
[79] LeBaron and Lister, 'Benchmarking global supply chains'.
[80] This section draws on Hans Krause Hansen and Tony Porter, 'What do numbers do in transnational governance?', *International Political Sociology*, 6:4 (2012), pp. 409–26.
[81] Fishman, 'The Wal-Mart you don't know', *Fast Company* (1 December 2003), available at: {http://www.fastcompany.com/magazine/77/walmart.html} accessed 3 July 2012.
[82] Thomas Wailgum, 'Will Wal-Mart suppliers see red at green edict?', *Computer World* (12 July 2009), and Jianfeng Wang, 'Economies of IT systems at Wal-Mart' – an historical perspective, *Academy of Information and Management Sciences Journal*, 9:1 (2006), pp. 45–66.
[83] Liz Parks, 'E-collaborative efforts: Partnerships fortify supply chain initiatives', *Drug Store News*, 23:2 (2001), pp. 1, 13.

collaboration than had been the case when Walmart was pioneering the technology. By 2011, the GCI's Global Upstream Supply Initiative (GUSI) had worked with GS1 standards to promote a standardised global model of close integration similar to the one that Walmart had developed, and to extend this back to the manufacturers' suppliers. The adoption of standardised GS1 communications technologies across the supply chain allows increased automation of procurement, material forecasting, inventory management, dispatch, receipt, consumption of materials and financial settlement, with all the efficiency and other performance benefits that accompany this.[84] VICS merged with GS1 US in 2012.[85] The benchmarking associated with the Global Scorecard displays far more publicness than was the case with the earlier period dominated by Walmart. However this publicness mostly involves collaboration among firms in improving their profitability despite some incorporation of other public concerns such as sustainability.

Comparing the cases

The importance of context is evident in both disaster risk reduction and supply chain management. Their effectiveness depends on sustaining flows of information and action that work with the benchmarks but involve far more than an interaction of the architects and target actors with a benchmarking mechanism. The networks that sustain the flows need to be embedded in local contexts, technical artefacts, other objects and routine practices. In the case of disaster risk reduction this can extend to the interactions of a subsistence farmer with soil and crops, and in the case of supply chains this can extend to the placement of an RFID tag in a package and the configuration of the corresponding software in each firm's digital systems that communicate with it.

The impetus towards expanding circles of engagement, which correspond to increased publicness, is evident in each case as well. In disaster risk reduction it is evident in the recognised importance of increasingly engaging civil society actors, citizens, and underlying risk factors, and in facilitating the circular flows of information that enhance the effectiveness of benchmarking. In supply chain management it is evident in the recognised importance for its effectiveness of engaging actors at all levels, in bringing in increasing numbers of actors, in promoting flows of information, and in continually reworking structures.

While the relative power of the architects and target actors and the character of the benchmark design are certainly significant in both cases, our ability to analyse and evaluate the benchmarking systems would be very limited if we were restricted to these variables. The performances of the benchmarks in the two cases are more closely related to the similar contextual factors that both face. In both cases, effective benchmarking requires a circular flow of information and action through networks that extend far beyond the benchmark itself, and engage with multiple actors in multiple locations. It still matters that the first case involves states and the second firms, since the degree of publicness is greater in the former case. However, the difficulty of establishing connections and soliciting engagement is similar across the

---

[84] GSI, 'Upsteam Integration Model (UIM)', available at: {http://www.gs1.org/upstream/uim} accessed 27 August 2014.

[85] GS1, 'GS1 US and Voluntary Interindustry Commerce Solutions Association to Merge', Press Release (10 September 2012), available at: {http://www.gs1us.org/about-gs1-us/media-center/press-releases/gs1-us-and-vics-merger} accessed 29 August 2014.

two cases, even though this primarily involved citizens in the first case and employees in the second.

The character of the design of the benchmarks in the two cases is important, but only when considered in relationship to the contexts with which they interact. For instance both the HFA and Global Scorecard benchmarks have online templates, and it is significant that the former automatically publicises the submissions while the latter carefully preserves the confidentiality of the information submitted by companies. This feature of the mechanism is related to the public character of the architects of the HFA process and the private character of the architects of the Global Scorecard. However not all benchmarking information involving states is made public, and not all benchmarking information involving firms is kept private. The larger significance of these differences relates to how the benchmark interacts with its contexts, which in the case of disaster risks cuts more directly across all organisational boundaries than is the case with supply chains.

It is not possible to generalise from these cases to claim that transnational benchmarking in general promotes publicness. In other cases self-interest and power asymmetries may be more important than in these cases. However these cases do confirm that benchmarking can provide an impetus towards greater publicness, as well as showing the importance of the contexts of benchmarking for assessing the relative influence of this impetus and the factors that work against it.

Conclusion

This article has argued that it is important to go beyond analysis of the architects, target actors, and design characteristics of a benchmark to analyse more systematically the contexts of benchmarking, and whether and how these sustain the networks carrying flows of information and action that are needed if a benchmark is going to be effective. It drew on actor-network theory in identifying a method for analysing contexts of benchmarking. It identified embeddedness and publicness as two the key aspects of contextual variation. It also argued that the logic of benchmarking created an impetus towards increased publicness, although this was constrained by the challenges associated with embeddedness and the self-interest of the actors involved.

The two cases demonstrated the importance of contexts for the effectiveness of benchmarking, and for assessing their capacity to solve collective global problems. The need for benchmarking to establish robust connections with local contexts, so that the flows of information and actions are embedded, was clear, whether this was subsistence farmers and their crops, or the RFID tagged products moving through the supply chain. While disaster risk reduction and supply chain management involve very different problems and actors, both displayed an impetus towards greater publicness over time, although this was uneven. The two cases were striking for their similarities, despite the differences in the actors involved and the problems they sought to address. This confirms the distinctiveness of benchmarking as a mode of governance, where the interactions between the benchmark and its contexts display certain similarities despite variations in the types of actors that constitute its architects and targets or how the benchmark itself is designed. While the two cases examined here were illustrative, there is much more valuable research which can be done on other cases of interaction between benchmarks and their contexts, more systematically identifying the conditions that contribute to the effectiveness and publicness of benchmarking.

Transnational benchmarking is inserted into an increasingly dense, complex, and accelerated global governance setting where it has to interact with a diverse set of actors spread out over multiple locations if it is to be effective. Rather than a measuring instrument, or a tool used by one set of actors in their interaction with another set, benchmarking, in its most distinctive form, is best conceptualised as a partially autonomous relay and accelerator in a circular and reflexive flow of actions and information. Benchmarking can be a collaborative mode of governance that can be used by a set of actors to solve their own collective problems. However, benchmarking can also still be used by powerful actors to advance their particular interests. The balance between this and publicness in any particular benchmarking system can only be assessed by analysing the way that the benchmark interacts with its contexts, and whether or how circular and reflexive networks carry information and actions from the benchmark through those contexts and then back again. There are many opportunities to further link the analysis in this article to related or complementary literatures and approaches, many of which have been noted above, including the literatures on experimentalist governance, transnational communities, global publics, global governmentality, and transnational value chains, among others.

How activists use benchmarks: Reformist and revolutionary benchmarks for global economic justice

LEONARD SEABROOKE AND DUNCAN WIGAN*

Abstract. Non-governmental organisations use benchmarks as a form of symbolic violence to place political pressure on firms, states, and international organisations. The development of benchmarks requires three elements: (1) salience, that the community of concern is aware of the issue and views it as important; (2) will, that activists and issue entrepreneurs will carry the message forward; and (3) expertise, that benchmarks created can be defended as accurate representations of what is happening on the issue of concern. We contrast two types of benchmarking cycles where salience, will, and expertise are put to the test. The first is a reformist benchmarking cycle where organisations defer to experts to create a benchmark that conforms with the broader system of politico-economic norms. The second is a revolutionary benchmarking cycle driven by expert-activists that seek to contest strong vested interests and challenge established politico-economic norms. Differentiating these cycles provides insights into how activists work through organisations and with expert networks, as well as how campaigns on complex economic issues can be mounted and sustained.

Leonard Seabrooke is Professor of International Political Economy and Economic Sociology in the Department of Business and Politics at the Copenhagen Business School, and Research Professor in the Department of International Politics at the Norwegian Institute of International Affairs. He is co-author of *Global Tax Battles* (Oxford: Oxford University Press, 2016). Seabrooke's work for this article was funded by the 'Professions in International Political Economies' project from the European Research Council (#263741-PIPES).

Duncan Wigan is Associate Professor of International Political Economy in the Department of Business and Politics at the Copenhagen Business School. He is co-author of *Global Tax Battles* (Oxford: Oxford University Press, 2016). Wigan's work for this article was funded by the 'Global Re-ordering: Evolution through European Networks' projected funded by the European Commission Framework Seven Programme (#266809-GR:EEN).

Introduction

Benchmarks are weapons of symbolic violence.[1] They are a form of domination from an implied consensus on how to rank and order, to assess who and what is worthy, to inform ways of being. As symbolic weapons benchmarks are useful for both reformist and revolutionary agendas, for those who use benchmarks to adjust behaviour within

[1] Pierre Bourdieu, *The Logic of Practice,* trans. Richard Nice (Stanford: Stanford University Press, 1990); Pierre Bourdieu, *Masculine Domination*, trans. Richard Nice (Stanford: Stanford University Press, 2001).

conforming norms, as well as those who seek to provide a direct challenge to how we think about the world and how it should be. Benchmarks are available to a range of actors nowadays. Weapons of symbolic violence in the twentieth century were primarily controlled by the state through the organisation of ideational and material resources around official certification systems, as well as the establishment of jurisdictions of competence in cooperation with professional associations.[2] Struggles over symbolic violence include the official consensus from the state and those who support it, insults from individuals expressing their views, and expert critics who could claim authority to their viewpoints.[3] In the postwar period the rise of private capital and of international civil society led to benchmarks being increasing used as a challenge to state-led forms of certification and standard-setting. Firms began using benchmarks as technical tools in the 1950s, to identify weaknesses in their systems. These benchmarks transformed into forms of managerial strategy in the 1980s and then 'calculative practices' to focus our minds on competitiveness and performance in the contemporary period.[4] Now the use of benchmarks and 'scorecards' is common practice, with bodies such as the International Organization for Standardization producing benchmarks and standards to guide corporate, governmental, and civil society behaviour.[5] Non-governmental organisations (NGOs) and expert-activists also put benchmarks forward, seeking to influence behaviour by providing their own systems of evaluation and judgement. Be they used by firms, NGOs, or public agencies, benchmarks are primarily 'reactive measures' that seek to induce conformity towards a value system that is commonly represented as a measuring tool that reflects an otherwise unseen ordering of units within a system.[6] Occasionally benchmarks can also challenge us to change how we see the units.

In their introduction to this Special Issue, André Broome and Joel Quirk discuss four constitutive elements of benchmarks, which include: (1) the reification of normative standards into observable categories; (2) commensuration through homogenising entities into a common metric; (3) symbolic judgments that create focal points for action; and (4) reactions from those targeted by benchmarks.[7]

---

[2] Andrew Abbott, *The System of Professions: An Essay on the Division of Expert Labor* (Chicago: University of Chicago Press, 1988).

[3] Pierre Bourdieu, 'The social space and the genesis of groups', *Theory and Society*, 14:6 (1985), pp. 723–4 (p. 732).

[4] As argued by Wendy Larner and Richard Le Heron, 'Global benchmarking: Participating "at a distance" in the globalizing economy', in Wendy Larner and William Walters (eds), *Global Governmentality: Governing International Spaces* (London. Routledge, 2004), pp. 212–32. See also, Tore Fougner, 'Neoliberal governance of states: the role of competitiveness indexing and country benchmarking', *Millennium: Journal of International Studies*, 37:2 (2008), pp. 303–26.

[5] Winton Higgins and Kristina Tamm Hallström, 'Standardization, globalization and rationalities of government', *Organization*, 14:5 (2007), pp. 685–704. See also, Hans Krause Hansen and Dorte Salskov-Iversen, 'Remodeling the transnational political realm: Partnerships, benchmarking schemes and the digitalization of governance', *Alternatives: Global, Local, Political*, 30:2 (2005), pp. 141–64, as well as Peter Gibbon and Lasse Folke Henriksen, 'A standard fit for neoliberalism', *Comparative Studies in Society and History*, 52:2 (2012), pp. 275–307. Benchmarks are commonly put forward by private actors who can make claims to expert authority, such as in areas like international accounting. See Tim Büthe and Walter Mattli, *The New Global Rulers: The Privatization of Regulation in the World Economy* (Princeton: Princeton University Press, 2011). An excellent case of how accounting standards are actually implemented can be found in Matthias Thiemann, 'The impact of meta-standardization upon standards convergence: the case of the international accounting standard for off-balance-sheet financing', *Business & Politics* (2013), DOI 10.1515/bap-2012-0011.

[6] Wendy Espeland and Michael Sauder, 'Rankings and reactivity: How public measures recreate social worlds', *American Journal of Sociology*, 113:1 (2007), pp. 1–40.

[7] André Broome and Joel Quirk, 'Governing the world at a distance: the practice of global benchmarking', *Review of International Studies*, 41:5 (2015), pp. 819–41.

We suggest that reification, commensuration, focal points, and reactions are all elements of what Pierre Bourdieu referred to as 'symbolic violence' and associated with gentle forms of control which define reality and include a system of domination in which those dominated are active or unwittingly complicit.[8] From this viewpoint, symbolic violence has risen as the 'most economical mode of domination because it best corresponds to the economy of the system'.[9] We argue that this is indeed the case when it comes to benchmarks for global economic justice. Benchmarks seek to disenchant the observer and convey good science; they entail a process of accumulating economic capital through symbolic capital, trust in a benchmark, that has redistributive consequences for societies.[10] Ultimately benchmarking seeks to redistribute claims to material and symbolic forms of capital.

This article is interested in how NGOs and expert-activists coordinate and create benchmarks. We seek to highlight how benchmarks for global economic justice differ in how they are organised, how the benchmarks are articulated, and the process through which they are tested for three key elements: *salience*, *will*, and *expertise*. We place great emphasis on how salience, will, and expertise are important for benchmarking cycles that can change transnational norms. Salience is a key test of relevance for NGOs and expert-activists, both from stakeholders they seek to persuade as well as from a political audience that can pay attention and change practices through hard and soft forms of governance.[11] Will is the initial condition for NGOs or expert-activists to seek benchmark creation, and as a benchmark develops the original spur must be transformed into political will to move from evaluation to enactment. Expertise is a key ingredient because often NGOs select issues on which they are not fully equipped to analyse, and because expertise must be recognised by peers lest those creating the benchmarks be viewed as flawed and inadequate or simply barking at the moon. For benchmarking to be successful we suggest that the trinity of will, salience, and expertise must be present, since failing a salience of expertise test sends those propagating the benchmark back to square one, or to develop a benchmark that will attract little attention or policy traction.

Benchmarks align with dominant thinking about how things should be but can also challenge the normative order. As such we differentiate *reformist* and *revolutionary* benchmarks. Reformist benchmarks seek to create change within the dominant system. The rise of benchmarks to assess accountability and corruption are an obvious case of reformist benchmarking, where NGOs are adding new practices that flow in the same direction as well established normative trends.[12] Revolutionary benchmarks seek to challenge how the basic units of the system, and the logic of the system are perceived. These benchmarks are often driven by expert-activists and must persuade NGOs, stakeholders, and political authorities that they have a point. Revolutionary benchmarks, as we clarify below, are frequently challenged for their expertise and salience, especially on methodological grounds where frictions can

---

[8] This concept is often considered the very core of Bourdieu's approach, see Mustafa Emirbayer, 'Tilly and Bourdieu', *The American Sociologist*, 41:4 (2010), pp. 400–22 (p. 412).
[9] Bourdieu, *The Logic of Practice*, p. 127.
[10] Ibid., pp. 133–4.
[11] See, in particular, Wendy A. Wong, *Internal Affairs: How the Structure of NGOs Transforms Human Rights* (Ithaca, NY: Cornell University Press, 2012).
[12] See, for example, the Geneva-based Human Accountability Partnership benchmarks for assessing accountability in humanitarian crises: Human Accountability Partnership, 'The 2010 HAP Standard in Accountablity and Quality Management' (Geneva: HAP, 2010), available at: {http://www.hapinterna-tional.org/pool/files/2010-hap-standard-in-accountability.pdf}.

easily emerge over good science rather than over politics. Claims to originate revolutionary benchmarks may be made from experience, from 'bearing witness' to injustices, but they also have to be supported by scientific expertise.[13] We suggest that reformist and revolutionary benchmarking cycles can be understood as ideal types, that is, as abstract heuristic devices that help us to understand a range of empirical cases and their differences.

Our specification of reformist and revolutionary benchmarks seeks to contribute to the extensive literature on how NGOs select issues to campaign on and how they muster ideational and material resources to push ahead. Our cases have been selected via inductive reasoning. In short, the mechanisms presented in the revolutionary case were presented to us over multiple interviews between 2011 and 2014 with tax justice campaigners and policy stakeholders. We then compared these dynamics with dominant themes in the International Relations literature and interviewed NGO professionals – those not tied to particular issues but generalists – about how they formulate benchmarks. These interviews were conducted in Brussels, London, and New York in 2012 to 2014. This piece focuses on the varied characteristics and mechanisms of benchmark entrepreneurship in the NGO world.[14] We seek to highlight how NGOs and expert-activists operate in a highly contested field of power where there are frictions between themselves and other types of organisations and groups, such as firms, states, and international organisations. Significant tensions exist within the organisations and groups as all seek a stake in the game.[15] Earlier work in this area highlighted how NGOs and transnational advocacy networks engaged in 'information politics' to get their messages across, and how they influenced states and international organisations by imposing reputation costs and shaming. To do so they selected issues linked to bodily harm and unequal access to rights, issues on which they could draw a large audience.[16] Should all this go well then it would be possible to establish a norm tipping point where transnational norms would be transformed.[17]

More recent research has built on these earlier insights to show how NGOs engage in gatekeeping to only select issues they feel are salient,[18] with managers within NGOs carefully vetting issues for their attributes prior to committing resources to them to ensure that salience is present. The competitive rather than cooperative side of NGOs has also been explored at length, with excellent scholarship on how NGOs compete with each other for resources to the detriment of those they seek to help,[19] as well as

---

[13] On claims to authority see Ole Jacob Sending, *The Politics of Expertise: Competing for Authority in Global Governance* (Ann Arbor: University of Michigan Press, 2015). On scientific expertise and career experience see Lasse Folke Henriksen, 'The global network of biofuel sustainability standard-setters', *Environmental Politics*, 24:1 (2015), pp. 115–37.

[14] Vivien Collingwood, 'Non-governmental organisations, power and legitimacy in international society', *Review of International Studies*, 32:3 (2006), pp. 439–54.

[15] Pierre Bourdieu, *The State Nobility: Elite Schools in the Field of Power* (Cambridge: Polity Press, 1996). Applied to international political economy see, in particular, Matthew Eagleton-Pierce, *Symbolic Power in the World Trade Organization* (Oxford: Oxford University Press, 2013).

[16] Margaret E. Keck and Kathryn Sikkink, *Activists Beyond Borders: Advocacy Networks in International Politics* (Ithaca, NY: Cornell University Press, 1998).

[17] Martha Finnemore and Kathryn Sikkink, 'International norm dynamics and political change', *International Organization*, 52:4 (1998), pp. 887–917.

[18] Wendy A. Wong, *Internal Affairs*; see also R. Charli Carpenter, 'Setting the advocacy agenda: Theorizing issue emergence and nonemergence in transnational advocacy networks', *International Studies Quarterly*, 51:1 (2007), pp. 99–120.

[19] Alexander Cooley and Ron James, 'The NGO scramble: Organizational insecurity and the political economy of transnational action', *International Security*, 27:1 (2002), pp. 5–39; Shareen Hertel, *Unexpected Power: Conflict and Change among Transnational Activists* (Ithaca, NY: Cornell University Press, 2007).

how NGOs often act as coalitions and interest groups that display distinct national differences.[20] Extensive studies have also pointed out how NGOs have become professionalised, bureaucratised, and even corporatised in how they mount and operate campaigns.[21] From this perspective, NGOs and benchmark creation may be easy due to greater in-house capacity and expertise, or may also be much more difficult because many NGOs have become lumbering bureaucratic machines built to campaign rather than house expertise.

Rather the viewing the increasing professionalisation and corporatisation of NGOs as a rational process through which they seek to attain influence within the dominant normative order of our time, we argue that benchmarks are double-faced and volatile, depending on who is advocating them.[22] Reformist and revolutionary types of benchmarks contain different characteristics of salience, will, and expertise that change the stages in a benchmarking cycle and the conditions of possibility. Reformist benchmarking, we argue, commonly relies on NGOs outsourcing to experts who then seek to create stakeholder buy-in on the way to creating benchmarks, which are then conveyed to political authorities. Because these benchmarks tend to align with public and private sector norms reformist benchmarks have a much greater chance of passing the test of political engagement and leading to what Sidney Tarrow refers to as 'certification', when authorities validate claims and performances made by actors.[23] By contrast, revolutionary benchmarks do not come from larger NGOs, whose managers choose not to approach issues in a way that damages their membership and resource base. Rather, revolutionary benchmarks emerge from expert-activists that then seek NGO buy-in. These expert-activists often directly challenge how we see units in the system and the logic of the system. As such their benchmarks come up against 'mature field' opposition from NGOs and corporate interests that are well established and question the expertise behind, and salience of, the benchmark.[24] Given this conflict political authorities are less likely to certify these benchmarks. Still, expert-activists have the potential to use benchmarks as weapons in a fight over what modes of domination prevail.[25] Benchmarks are important weapons in global battles over the right to govern economic activity.[26]

The article is structured as follows: first we provide the Reformist Benchmarking Cycle, establishing how it works and providing two examples of benchmarks through Transparency International's Corruption Perceptions Index and the Gold Standard's development of certification credits based on the benchmarking of renewable energy

[20] Elizabeth A. Bloodgood, 'The interest group analogy: International non-governmental advocacy organizations in international politics', *Review of International Studies*, 37:1 (2011), pp. 1–28; Helen Yanacopulos, 'The strategies that bind: NGO coalitions and their influence', *Global Networks*, 5:1 (2005), pp. 93–110. See also Sarah S. Stroup, *Borders Among Activists: International NGOs in the United States, Britain, and France* (Ithaca, NY: Cornell University Press, 2012).

[21] Stephen Hopgood, *Keepers of the Flame: Understanding Amnesty International* (Ithaca, NY: Cornell University Press, 2006); Peter Dauvergne and Genevieve LeBaron, *Protest Inc: The Corporatization of Activism* (Cambridge, UK: Polity, 2014).

[22] Pierre Bourdieu, *The Rules of Art: The Genesis and Structure of the Literary Field,* trans. Susan Emanuel (Stanford: Stanford University Press, 1995), p. 358, fn. 14.

[23] Sidney Tarrow, *The New Transnational Activism* (Cambridge: Cambridge University Press, 2005), p. 194.

[24] Royston Greenwood and Roy Suddaby, 'Institutional entrepreneurship in mature fields: the big five accounting firms', *The Academy of Management Journal*, 49:1 (2006), pp. 27–48.

[25] See, in particular, Roy Suddaby and Thierry Viale, 'Professionals and field-level change: Institutional work and the professional project', *Current Sociology*, 59:4 (2011), pp. 423–42.

[26] Leonard Seabrooke and Duncan Wigan, *Global Tax Battles: The Fight to Govern Corporate and Elite Wealth* (Oxford: Oxford University Press, 2016).

or energy efficiency projects. Both benchmarks conform to the normative order of our time and seek to reform the treatment of an issue by challenging core assumptions about corruption or environmental markets. Both benchmarks were also developed as offshoots from larger organisations. This leads us to the second section, where we provide the Revolutionary Benchmarking Cycle, establishing how it differs from the previous cycle. We provide an example of the Tax Justice Network's Financial Secrecy Index (FSI), a benchmark that directly challenges assumptions about how core advanced industrialised states operate and the logic of how 'tax havens' operate in the international political economy. We highlight how expert-activists working through the Tax Justice Network have been key to the FSI and challenged by a 'mature field' in accountancy and politics, primarily on 'non-political' methodological grounds. Evidence for these cases has been gathered through extensive interviews with key stakeholders, as well as document analysis from public sources or private sources obtained from interviewees. In the final section we discuss the two benchmarking cycles and why it is important to differentiate reformist and revolutionary types.

Reformist benchmarking cycles

Reformist benchmarking cycles seek to push forward progressive agendas within the established organisational frameworks and normative order of the day. Reformist benchmarking is particularly useful to larger NGOs who concentrate on mounting and implementing campaigns on their selected issues. These NGOs, such as Oxfam, are highly bureaucratised and professionalised, with staff selected for their professionalism as much as for their passion in seeking to push forward a particular issue. As noted in the literature, such large NGOs are abundant with 'management speak' and a key concern for them is how to maintain lines of revenue and membership through the careful selection of issues.[27] Such agenda vetting is important for NGOs when choosing issues since those that are too complicated or carry mixed messages are unlikely to play well.[28] Issue selection is also guided by the industrial structure surrounding it,[29] with some issues easier to campaign on than others – compare, for example, 'fair trade' campaigns for coffee and chocolate and sex slavery in large hotel chains.[30] Since the fair trade mark was awarded in the Netherlands in 1988, NGOs supporting it have had a long history of cooperation with firms to provide a combination of salience, will, and expertise, especially in the identification of global value chains.[31] By contrast, initiatives such as the Global Business Coalition Against Human Trafficking have to make significant strides to combine their campaigns with the commercial interests and reputations of hotel chains in which sex slavery is an informal part of the

---

[27] Stephen Hopgood, *Keepers of the Flame*, pp. 133–4.
[28] R. Charli Carpenter, 'Vetting the advocacy agenda: Networks, centrality and the paradox of weapons norms', *International Organization*, 65:1 (2011), pp. 69–102.
[29] Rachel Schurman, 'Fighting frankenfoods: Industry structures and the efficacy of the anti-biotech movement in Western Europe', *Social Problems*, 51:2 (2004), pp. 243–68.
[30] Laura Raynolds, 'Mainstreaming fair trade coffee: From partnership to traceability', *World Development*, 37:6 (2009), pp. 1083–93; Cordine Gendron, Véronique Bisaillon, and Ana Isabel Otero Rance, 'The institutionalization of fair trade: More than just a degraded form of social action', *Journal of Business Ethics,* 86:1 (2009), pp. 63–79.
[31] Anna Hutchens, *Changing Big Business: The Globalization of the Fair Trade Movement* (Cheltenham: Edward Elgar, 2009).

supply chain.[32] The selection of issues will constrain how organisations can campaign and what institutional repertoires they have available to them.[33]

Larger NGOs also have an interest in creating 'scale free' networks that have low costs for new members to join so that expansion and communication on issues is made more efficient.[34] Choosing issues with obvious political and moral salience is the key way to constructing these networks, as is involving key stakeholders in testing what counts as a salient issue. For these reasons, many NGOs cooperate closely with firms and non-civil society actors to harness greater support and pass salience tests.[35] Such ties provide access to those who may have the expertise that is missing from the NGO, and robust networks assist in the creation of benchmarks that can also gather political will. In short, there are many incentives for NGOs that are interested in creating benchmarks to conform to existing normative standards and then propose the strengthening of a particular element rather than advocating a revolutionary change in how the system is perceived.

Figure 1, below, presents our view of a reformist benchmarking cycle. To walk us through the various steps in the cycle, at the twelve o'clock position is the transnational norm which an organisation or activists seek to modify or change through the creation of a benchmark. This is an initial test of will that, if present, leads to the selection of an issue by the organisation at Step 1. As discussed, this process is carefully managed within NGOs and subject to a range of perceptions about what has political and moral salience and what does not. Following issue selection the organisation then moves to expert engagement that can take place within the organisation, outside the organisation (through external consulting), or through the establishment of an external expert group by the primary organisation. This is Stage 2. The next step is to engage stakeholders once the organisation has information from the experts. This is the first test of salience and stakeholders, be they other NGOs, firms, state agencies, or international organisations, can provide feedback on what will work and will not work, perhaps leading to an early death to this initiative. If the salience test at Step 3 fails then it is back to Step 1. However, if successful the next step is benchmark creation, where experts provide their expertise alongside the organisation's normative reformist goals. This is Step 4 where the information from expert engagement is directly put to use. The next stage is political engagement where there is a second test of salience as well as a test of expertise. Political authorities may question the expertise of those involved at Step 2, especially when the issue is politically and economically sensitive, sending the benchmarking process back to the expert engagement stage.[36] Political authorities may also discount the benchmarking initiative as unimportant altogether, pushing it back to Step 1 where the organisation

---

[32] Global Business Coalition Against Human Trafficking, available at: {www.gbcat.org} accessed 10 October 2014.

[33] Issue selection follows who can control issues in professional and organisational networks, see: Lasse Folke Henriksen and Leonard Seabrooke, 'Transnational organizing: Issue professionals in environmental sustainability networks' (2016), *Organization*, DOI: 10.1177/1350508415609140.

[34] Wendy Wong, *Internal Affairs*.

[35] Philipp Pattberg, 'The institutionalization of private governance: How business and nonprofit organizations agree on transnational rules', *Governance*, 18:4 (2005), pp. 589–610. See also, Luc Fransen, *Corporate Social Responsibility and Global* Labour *Standards: Firms and Activists in the Making of Private Regulation* (Abingdon: Routledge, 2012).

[36] This can also occur via the formation of expert groups that reject benchmarks that challenge the conventional terms of debate, see Leonard Seabrooke and Eleni Tsingou, 'Distinctions, affiliations, and professional knowledge in financial reform expert groups', *Journal of European Public Policy*, 21:3 (2014), pp. 389–407.
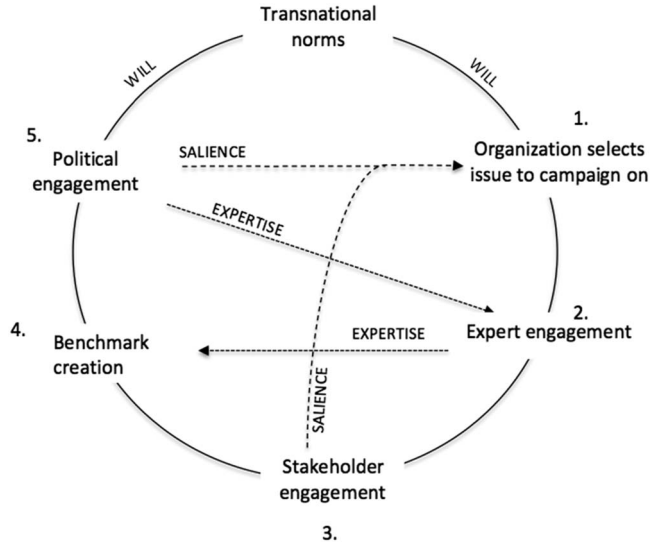
Figure 1. *Reformist benchmarking cycle*

can vet and select the issue once more. However, if Step 5 is passed and there is a combination of salience, expertise, and also political will then the transnational norm can be transformed. We present this cycle as an ideal type that, as with all ideal types, should be broken down and reassembled once it fails to adequately depict reality. We also note that variation in benchmarking cycles can also be understood as emerging from the differences in the articulation of conforming and contesting norms for reformist and revolutionary issues at different steps in the cycle.

Reformist benchmarking cycles are the most common form given that organisations can select issues that are likely to have salience, where expertise can be found due to the organisational and commercial benefits associated with established systems, and where will is more readily mustered as it already aligns with present norms. As such there are many examples to draw upon, but we highlight two prominent cases of reformist benchmarking that follows the cycle described above: Transparency International's Corruptions Perceptions Index (CPI), and the World Wildlife Fund for Nature's support for the development of the Gold Standard as certifiers of sustainability within carbon markets.

The Corruption Perceptions Index (CPI) was launched in the mid-1990s by Transparency International (TI). TI was established in 1993 by Peter Eigen who, frustrated with the impact of corruption on projects in Africa, retired from the World Bank to combat corruption. TI set out as an activist organisation led from the top.[37] Founding members included officials of international organisations, former ministers from developing countries, and elite businessmen who chose corruption as an issue to work on.[38] TI has more than 100 National Chapters who engage in domestic

---

[37] Ellen Gutterman, 'The legitimacy of transnational NGOs: Lessons from the experience of transparency international in France and Germany', *Review of International Studies*, 40:2 (2014), pp. 391–418.

[38] For example, former US President Jimmy Carter joined the advisory board in 1998. More generally, see Peter Larmour, 'Civilizing techniques: Transparency international and the spread of anti-corruption', in Brett Bowden and Leonard Seabrooke (eds), *Global Standards of Market Civilization* (London: Routledge, 2006), pp. 95–106.

anti-corruption initiatives, while the international secretariat is located in Berlin. TI flagship projects include its 'Sourcebook', which was launched in 1996 to provide a step-by-step guide to combating corruption in different sectors. This works in conjunction with a best practice oriented 'National Integrity System', which provides a qualitative assessment of a country's institutional capacity to provide good governance. By far the most visible TI product is the CPI launched in 1995.

The CPI was developed in an environment of increased attention on reforming polities in the post-Cold War period and the diffusion of democracy.[39] As suggested by Honghing Wang and James T. Rosenau, TI's formation of a CPI would have 'a better chance of success if it does not contradict the interest of powerful actors'.[40] In 1994, the OECD adopted a recommendation to criminalise foreign bribery and in 1997 its then 34 member states agreed to criminalise the paying of bribes to officials in other countries.[41] TI's proliferation of anti-corruption norms was emboldened by ready support from multinational firms, powerful states, and international organisations who had a pre-existing interest in reducing corruption.[42]

The CPI ranks countries and territories on a scale of 1 to 100 according to perceived levels of corruption among public officials and politicians. A score of 0 reflects very high levels of corruption. The CPI is composite index drawing from polls and surveys by political risk analysis firms, business schools, civil society organisations, and international organisations including the African Development Bank, the World Bank, and the Economist Intelligence Unit.[43] TI's Berlin Secretariat does not make a strong claim to expertise in the fields of law, policing or justice, with professional skills in the promotion and dissemination of ideas. Rather, for TI 'Expertise when needed was drawn in *ad hoc*.'[44] Index creation relies on the engagement of expertise embedded in existing analysis by a host of actors already promoting an anti-corruption norm. The index places emphasis on institutional quality, public officials' proclivity to demand bribes, and the dependence of business performance on corrupt transactions. The CPI does not provide a full picture of corruption in a given jurisdiction, but focuses on political and administrative corruption.

In the last decade the CPI has been incorporated into governance measures by the World Bank, International Monetary Fund, OECD, and the Bank for International Settlements, which speaks to the 'certification' of the benchmark at the heart of global governance.[45] Improving CPI scores are also considered to be

[39] Jennifer L. McCoy and Heather Heckel, 'The emergence of a global anti-corruption norm', *International Politics*, 38:1 (2001), pp. 65–90.

[40] Hongying Wang and James T. Rosenau, 'Transparency international and corruption as an issue of global governance', *Global Governance*, 7:1 (2001), pp. 25–49 (p. 41).

[41] OECD Convention on Combating Bribery of Foreign Public Officials in International Business Transactions (Paris, OECD, 2011).

[42] See also Hun Joon Kim and J. C. Sharman, 'Accounts and accountability: Corruption, human rights, and individual accountability norms', *International Organization*, 68:2 (2014), pp. 417–48.

[43] Transparency International, 'Corruption Perceptions Index 2013: In Detail', available at: {http://www.transparency.org/cpi2013/in_detail}.

[44] Larmour, 'Civilizing techniques', p. 104.

[45] In 2008, the World Bank's 'Control of Corruption' indicator was based on the 11 sources used by TI in the same year and an additional 14 based on the perceptions of experts and opinion polls. See Vito Tanzi, 'Corruption Around the World: Causes, Consequences, Scope and Cures', International Monetary Fund, Department of Fiscal Affairs, WP98/63 (Washington, IMF, 1998); Anja Rohwer, 'Measuring corruption: a comparison between the transparency international's corruption perceptions index and the World Bank's worldwide governance indicators', *CESifo DICE Report*, 7:3 (2009), pp. 42–52. For the WGI, see {http://info.worldbank.org/governance/wgi/index.aspx#home}.

crucial in who receives overseas development assistance.[46] Yet, the index fails to capture the supply side of corruption and the culpability of private actors.[47] The 2013 CPI ranked 177 countries with Denmark heading the list with a score of 91 and Afghanistan, North Korea, and Somalia at the bottom with scores of 8. Countries deemed to be performing in the top twenty of corruption perceptions in 2013 included Switzerland, Netherlands, Singapore, Luxembourg, Hong Kong, and Barbados. These countries fare less well in the revolutionary benchmark we present below. Benchmarking issues of global economic justice for reformist purposes will often generate broad scope and traction, finding function in ready requirements.

A second benchmark that conforms to the reformist cycle is the 2003 Gold Standard initiated in the wake of the 1997 Kyoto Protocol by the World Wildlife Fund (WWF) to remedy the failure of Clean Development Mechanism (CDM) and Joint Implementation (JI) projects to deliver 'additionality' and genuine added environmental and social benefits. The WWF encouraged the initiative to become an independent body having realised the market 'would be better served by an independent standard with really tough experts serving it'.[48] The Gold Standard exceeds the requirements of the UN's CDM Executive Board and generates credits representing one tonne of reduced carbon dioxide equivalent and carrying a price premium in the voluntary market. The development of the standard relied on the work of a Technical Advisory Committee consisting of experts from government, civil society organisations, and firms. At the outset the advisory committee outnumbered Foundation employees by four to one. In an interview with the CEO it was noted that expert consultation helps reconcile potentially divergent market and non-market goals to augment legitimacy.[49] The benchmark emerged from a two-year consultation process, where stakeholder involvement was gathered from a large section of the NGO community. Without this support, and with the Standard now formally endorsed by over eighty NGOs, the value of certified carbon credit was at risk. According to the CEO:

We derive credibility and brand value from operating on behalf of CSOs [civil society organisations] and always held the view that the Gold Standard needed the support of many NGOs. There is real value associated with endorsements from a wide variety of NGOs. For a corporate investor this lowers risk, that NGOs will not turn around and beat you up about it, your CSR [corporate social responsibility].[50]

A first round of consultations with members of environmental groups and representatives of other sectors were held before consultants with experience in project development and verification drafted a project assessment framework. There were also consultations with representatives of private sector firms active in the market for greenhouse gas allowances and government officials responsible for this area in South Asia, Philippines, Japan, South Africa, the United States, Western and

[46] Omar Hawthorne, 'Transparency International's Corruption Perceptions Index: "best flawed" measure on Corruption?' (unpublished mimeo, 2013), available at: {http://campus.hec.fr/global-transparency/wp-content/uploads/2013/10/Hawthorne-HEC.pdf}.

[47] For a critique of the CPI see Alex Cobham, 'Corrupting perceptions: Why transparency international's flagship corruption index falls short', *Foreign Policy* (22 July 2013).

[48] Jennifer Morgan, former head of WWF's climate change programme quoted in 'As The Gold Standard celebrates its 10th anniversary, it is expanding its reach beyond carbon', available at: {http://www.goldstandard.org/10-years-of-gold-standard-2}.

[49] Interview Gold Standard Foundation CEO.

[50] Ibid.

Eastern Europe.[51] By 2014, the Gold Standard had certified more than 1,000 projects and taken 20 million tonnes of carbon out of the atmosphere. Now it is a recognised benchmark in the carbon commodity markets, with large firms supporting its development and specialist consultancies continuing to provide certification and technical expertise to improve further reformist cycles.[52] Revolutionary benchmarking cycles face greater challenges.

Revolutionary benchmarking cycles

Revolutionary benchmarking cycles seek to interpret how we view the system surrounding an issue, as well as the character of units within the system. They differ from reformist benchmarking cycles in that tests of salience, will, and expertise are tougher, confronting vested interests who have clear reasons not to overturn their understanding of how the system works. An issue for revolutionary benchmarking is that successful benchmarks often seek to remove much of the moral and political tension inherent in the benchmarking process. As noted above, the editors for this Special Issue note that benchmarks involve reification, commensuration, focal points, and reactions. Much reification and commensuration is about rendering technical, political, and moral complexity into easily read metrics and indices.[53] Reformist benchmarking cycles align interests between NGOs, experts, and those concerned in the public and private sectors, allowing reification and commensuration to be effective. This is quite different for revolutionary benchmarking cycles, where expertise is probed for flaws in methodological construction.

Figure 2, below, presents our view of a revolutionary benchmarking cycle. As with the reformist benchmarking cycle we have five steps, but the character of the steps in the revolutionary benchmarking cycle are quite different. At Step 1 expert-activists select an issue to campaign on. This selection is based more on expertise and transformation of moral intent than the typical gatekeeper function found in reformist benchmarking cycles, where key managers inside the NGO carefully vet what can be campaigned upon and what cannot.[54] Step 2 is benchmark creation where the expert-activists develop the methodology for benchmarking and may consult with other experts to verify its soundness. Rather than seeking wider stakeholder engagement the expert-activists seek to have other NGOs 'buy-in' to the benchmark and support its validity. This is a test of salience since the NGOs need to buy into not only the issue as politically and socially important but also have to agree with, or defer to, the expert-activists on questions of data gathering and significance, and especially methodology. That is Step 3. At Step 4 we note mature field opposition, where interests associated with the established institutions being challenged will confront the expert-activists on their expertise and fitness to

[51] 'The Gold Standard – Background and Overview' (The Gold Standard, July 2003), available at: {http://assets.panda.org/downloads/thegoldstandardoverview.doc}.

[52] 'The Gold Standard Requirements' were supported by large firms active in the carbon credit market such as Firstclimate, Climatecare, and BNP Paribas and consultancies such as Ecofys, Tüv Südand Tricorona. Available at: {http://www.ecofys.com/files/files/gsv2_requirements_20080731_2.0.pdf}.

[53] Hans Krause Hansen and Arthur Mühlen-Schulte, 'The power of numbers in global governance', *Journal of International Relations and Development*, 15:4 (2012), pp. 455–65.

[54] On 'identity switching' among expert-activists, see Leonard Seabrooke and Duncan Wigan, 'Emergent Entrepreneurs in Transnational Advocacy Networks: Professional Mobilization in the Fight for Global Tax Justice', GR:EEN Working Paper No. 42, Centre for the Study of Globalisation and Regionalisation, University of Warwick (December 2013).
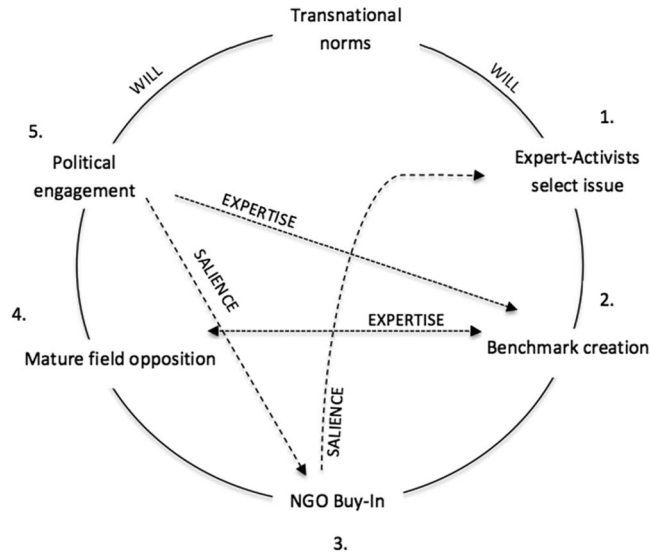
Figure 2. *Revolutionary benchmarking cycle*

judge and create a benchmark.[55] Given that mature field institutions have a lot of resources and established strategies to communicate with both the public and political authorities, opposition on the grounds of insufficient expertise will be treated seriously. The expert-activists will have to fight back to demonstrate that their benchmark and findings are legitimate and valid, leading to an interactive process. If there is enough momentum to carry on then, in Step 5, there is political engagement where political actors can also question the expertise within the benchmark. This is also a test of salience that is directed primarily at the NGOs who have bought into the benchmark rather than the expert-activists, since the NGOs are the established parties. If this test of salience fails then there can be a knock-on effect of NGOs reconsidering their own salience test, sending the expert-activists back to Step 1.

A revolutionary benchmarking cycle can be found in the Financial Secrecy Index (FSI). The FSI contests both the CPI and alternative lists of 'tax havens'. While the CPI focuses on public officials and politicians, the FSI highlights the issue of financial transparency as systemic. The expert-activists behind the FSI, the Tax Justice Network (TJN), initially sought cooperation with TI in 2008, but were unable to form an agreement on strategy. TJN representatives interpret the reluctance of TI to directly cooperate as due to American staff considering tax issues too difficult to communicate and not wishing to disturb the success of the CPI. A key member of TJN communicated to us that 'TI had CPI running, they had this good selling point, they were all over the world with the index once a year … they depend on it for organizational survival.'[56] The OECD move strongly on tax issues with the blacklisting of 35 jurisdictions in 2000 for 'harmful tax competition'.[57] The OECD furthered their efforts in 2002 with the promotion of Tax Information Exchange Agreements (TIEA) and their deployment in the wake of the global financial crisis has been viewed

---

[55] Royston Greenwood and Roy Suddaby, 'Institutional entrepreneurship in mature fields'.
[56] Interview, TJN representative, 2014.
[57] See J. C. Sharman, *Havens in a Storm: The Struggle for Global Tax Regulation* (Ithaca, NY: Cornell University Press, 2006).

sceptically, with commentators questioning the mechanism's substance.[58] TJN argue that the benchmarks provided by TI and the OECD were both ineffective.[59]

For the world of NGOs, figuring out what to do about 'tax havens' is an ongoing thorny issue. While Oxfam identified 'tax havens' as a problem in the late 1970s there was no coherent movement on the 'problem of offshore' until the 2000s when activists organised through TJN. Established in 2003, TJN is a key actor on tax justice issues and strongly influential in defining campaign issues, targets, and policy content for the wider NGO community. TJN operates in an increasingly crowded policy space where many other NGOs and international organisations compete, including the OECD, the IMF, and private organisations such as the International Accounting Standards Board. This issue area is a mature field where public and private interests are well established. NGOs such as ActionAid have sought to mount a tax campaign on 'Tax Power', launched to increase focus on corporate tax abuse and encourage the increasing transparency of corporations and governments. Christian Aid has subsumed its tax efforts under the 'IF campaign' coalition targeting the G8. These well-resourced NGOs would like to tackle tax issues but have not developed revolutionary benchmarks. ActionAid have discussed the need for a benchmark on taxation, noting that 'Benchmarking the responsibility of a tax policy's content will inevitably be more controversial than benchmarking the management of that policy.'[60] As a consequence, ActionAid's tax benchmarking is more of an investor guide to make them aware of risks associated with tax planning, a reformist agenda. In contrast to this, the FSI provides a revolutionary benchmark in seeking to reconfigure how the units in the system and the system as a whole are conceived.

The FSI places jurisdictions on an index based on the level of secrecy they provide and the scale of their activities in global financial markets.[61] It is built upon a qualitative component that evaluates a jurisdiction's laws, regulations and tax treaties, and a quantitative component, the Global Scale Weight, which attaches a weighting to each jurisdiction on the basis of its share in global financial services exports. As a result the conventional imaginary of 'tax havens' is challenged. The 2013 FSI ranks the United States 6th, Germany 8th, and Japan 10th for financial secrecy, implicating some of the largest economies and most powerful nations in tax abuse and the facilitation of illicit financial flows. As a benchmark the FSI seeks to expose the hypocrisy of the G8 and place pressure on these core states to reform, rather than seeking to displace blame on the largely small island economies, those understood to populate the offshore world. The influential 1981 Gordon Report questioned the use of the term 'tax haven' early on, noting that the:

term 'tax haven' has been loosely defined to include any country having a low or zero rate of tax on all or certain categories of income, and offering a certain level of banking or

---

[58] M. J. McIntyre, 'How to end the charade of information exchange', *Tax Notes International*, 56:4 (2009), pp. 255–68; Markus Meinzer, 'The Creeping Futility of the Global Forum's Peer Reviews', Tax Justice Briefing (London: TJN, 2012), available at: {www.taxjustice.net/cms/upload/GlobalForum2012-TJN-Briefing.pdf}.

[59] See Richard Eccleston, *The Dynamics of Global Economic Governance: The Financial Crisis, the OECD, and the Politics of International Tax Cooperation* (London: Edward Elgar, 2013); Ronen Palan and Duncan Wigan, 'Herding cats and taming tax havens: the US strategy of "not in my backyard"', *Global Policy*, 5:3 (2014), pp. 334–43.

[60] ActionAid, 'Tax Responsibility: An Investor Guide' (Chard: ActionAid, 2013), p. 2.

[61] See: {http://www.financialsecrecyindex.com/} for the 2013 FSI.

commercial secrecy. Applied literally, however, this definition would sweep in many industrialized countries not generally considered tax havens, including the United States.[62]

'Tax haven' has been a useful term precisely because it shifts blame away from the core of the OECD member states and prolongs a transnational normative position that Swiss private banking is acceptable while shell companies in the Caymans are not. Challenging this, the FSI is designed to draw in powerful states not usually associated with tax abuse, financial instability and illicit finance. Secrecy jurisdictions are defined as: 'A jurisdiction which provides facilities that enable people or entities to escape or undermine the laws, rules and regulations of other jurisdictions elsewhere, using secrecy as a prime tool.'[63] The FSI points to hypocrisy in tax treatment within the OECD, something that experimental work involving the deception of tax avoidance and evasion service providers has also revealed.[64]

For the FSI, the 15 'Key Financial Secrecy Indicators' (KFSI) gauge a jurisdiction on the basis of the capacity to automatically exchange information and name the beneficial owners of companies, the extent to which country-by-country reports are required, whether provisions against money laundering are in place, whether a jurisdiction's tax administration uses taxpayer identifiers for efficient administration, and if trusts with flee clauses are permitted. Equally weighted scores on the 15 KFSIs are combined to form a secrecy score, ranging for 0 to denote complete transparency and 100 to denote complete secrecy. In the 2013 FSI, Switzerland was ranked first with a secrecy score of 78 and a global scale weight of 4.916. Montserrat ranked last, at 82nd with a secrecy score of 74 and a global scale weight of 0.000. Notably, Nairu ranked 81st despite a higher secrecy score than Switzerland due to a global scale weight of 0. The dataset underlying the index is made available online and linked to data sources for verification.[65]

The index measures secrecy rather than demarcating tax havens from virtuous jurisdictions or 'offshore' from 'onshore'. TJN popularised the phrase 'secrecy jurisdiction' to more accurately capture the activities and institutions implicated in tax abuse and financial improbity. This linguistic move overcame the territorial implications of the term 'offshore'. Any jurisdiction can be a tax haven for residents of another country.[66] Consequently, the FSI places jurisdictions on a secrecy spectrum, rather than nominating jurisdictions compliant or noncompliant. Dariusz Wójcik notes 'tax havenry' or 'offshoreness' is a matter of degree, not a binary variable.[67] In replacing the term 'tax haven', the FSI highlights that the activities of 'secrecy jurisdictions' are designed not so much as to attract mobile capital into a jurisdiction, but to provide a legal basis upon which to claim that an activity occurs nowhere or elsewhere for any interested fiscal authorities. Placing jurisdictions on a

---

[62] Richard Gordon, 'Tax Havens and Their Use By United States Taxpayers - An Overview' (US Treasury: Washington, DC, 1981).

[63] Markus Meinzer, 'White Paper on the FSI 2011. Where to Draw the Line? Identifying secrecy jurisdictions for applied research' (London: Tax Justice Network), available at: {http://www.taxjustice.net/cms/upload/pdf/FSI_2012_Cut-Off-Point.pdf}.

[64] Michael G. Findley, Daniel L. Nielson, and J. C. Sharman, *Global Shell Games: Experiments in Transnational Relations, Crime, and Terrorism* (Cambridge: Cambridge University Press, 2013).

[65] Tax Justice Network, 'Financial Secrecy Index', available at: {http://www.financialsecrecyindex.com/introduction/methodandconcepts} accessed 10 September 2014.

[66] Sol Picciotto, *International Business Taxation: A Study in the Internationalization of Business Regulation* (Cambridge: Cambridge University Press, 1992).

[67] Dariusz Wojcik, 'Where governance fails: Advanced business services and the offshore world', *Progress in Human Geography*, 37:3 (2013), pp. 330–47.

secrecy spectrum also broadens implicated practices beyond tax to include issues of financial stability and illicit finance.[68]

Issue selection for expert-activist benchmarks is by the activists as opposed to gatekeepers within a NGO. Expert-activists must define an issue and then seek traction and uptake among governments, NGOs, regulators, and the media. This places considerable pressure on internal expertise and experience. The FSI was developed by a team within TJN with on the ground experience in accounting, corporate service provision, and law. While the reformist benchmarking cycle passes through stages of expert consultation and stakeholder engagement before benchmark creation, expert-activists select an issue and jump straight to benchmark creation. That expert engagement and stakeholder consultation is absent increases the benchmark's vulnerability to scrutiny on methodological grounds. NGOs need to accept that the issue is salient and that the expertise behind the benchmark is sufficient before buying-in to it. Experts in mature fields will be unwilling to engage with benchmark creation where the benchmark challenges their position and revenue.

Only in the wake of the benchmark being launched does it come under concerted scrutiny. Mature field actors who are directly threatened by benchmark will question the expertise used in its creation. Geoff Cook, CEO of Jersey Finance Limited, responded to the 2011 FSI by contrasting it with appraisals by the IMF and OECD, stating that:

Jersey's ranking in the Index is nonsensical. Jersey is one of the safest and best regulated international finance centres (IFC), as demonstrated by credible, independent assessments by internationally accepted organisations like the OECD and IMF.[69]

The FSI is rejected by those in mature fields, who draw upon the weapons of stronger players in the system. With the rise of tax issues to the top of the international policy agenda, the 2013 FSI attracted considerable opposition. Caymans Finance commissioned an analysis by Aaron Smallwood, an economist from the University of Texas-Arlington, to dismantle the FSI. The critique targeted both the equal weighting of the KFSIs and the index weighting based on relative size of financial service exports. Smallwood argues that the 15 KFSIs do not contribute equally to levels of financial secrecy. For instance, while automatic information exchange is thought to contribute more to financial transparency than engagement in the tax treaty network the FSI weights both indicators equally. In calculating the index he notes that the:

United States, which has the sixth highest aggregate score for FSI, has the 56th highest score for S[ecrecy] … the jurisdiction with the worst secrecy score, Samoa, actually ranks 76th in terms of TJN's total FSI, which is because of the small size of Samoa's financial industry.[70]

These interventions generate methodology battles that expert-activists are obliged to fight.[71] Opposition to revolutionary benchmarks also comes from international

[68] Leonard Seabrooke and Duncan Wigan, 'Global wealth chains in the international political economy', *Review of International Political Economy*, 21:1 (2014), pp. 257–63.
[69] Sean O'Hare, 'Jersey Slams Financial Secrecy Index as "nonsensical"', *The Telegraph* (4 October 2011), available at: {http://www.telegraph.co.uk/finance/personalfinance/expat-money/8806246/Jersey-slams-financial-secrecy-index-as-nonsensical.html}.
[70] Aaron Smallwood, 'An Analysis of the Efficacy of the Tax Justice Network's Methodology in Constructing a Secrecy Index' (Cayman Finance, 2014).
[71] For TJN's response to Dr Smallwood, see: {http://www.financialsecrecyindex.com/PDF/TJN2014_FSI-CaymanResponse_20-OCT-2014.pdf}.

organisations that are heavily invested in seeing issues in a particular way.[72] On the FSI, the Head of Tax Policy and Administration at the OECD, Pascal Saint-Amans, stated in 2012:

> I wish that their financial secrecy index would be a bit more robust. They have a secrecy index that built on a number of criteria that are disputable and based on information that is outdated … They came up with a financial secrecy level made up with a mathematical formula to get countries they want on the top of the list to be on top of the list. That's disingenuous, it's not fair, it's not good and it sends the wrong message.[73]

This attracted a direct rebuttal in the same journal from the Director of TJN, who argued that the OECD is unable to tackle financial secrecy issues.[74] Despite mature field opposition the FSI has gained political traction with private and public organisations.

The FSI has become a default tool for the broader NGO community now working on tax issues with ActionAid, Publish What You Pay, Christian Aid, and a host of NGOs all buying-in to the FSI in campaign materials. In 2013, the Centre for Global Development in the US began using the FSI as one component in the Commitment to Development Index, measuring the developmental value of the policies of 27 OECD members.[75] The Basel Anti-Money Laundering Index by the Basel Institute of Governance includes the FSI as one criterion in its assessment of money laundering risk, and has defended the FSI on its methodology.[76] The 2013 Bribery and Corruption Awareness Handbook for Tax Examiners and Tax Auditors mentions a high FSI score as an external environment indicator of a high risk country.

There has also been some interest from the private sector in applying the FSI. Two German sustainability rating agencies, imug and Oekom research, incorporated the FSI into their rating processes on bank bonds and country ratings.[77] The French bank Crédit Coopératif decided in 2012 not to maintain or establish affiliates, nor to invest or finance projects in non-European jurisdictions that received a secrecy score higher than 70, leading to the exclusion of 47 countries.[78] Interestingly, the FSI has also spurred an opposing benchmark, with the US-based Tax Foundation launching its International Tax Competitiveness Index in 2014. This benchmark 'measures the degree to which the 34 OECD countries' tax systems promote competitiveness

---

[72] On how international organisations make issues and members 'legible' see André Broome and Leonard Seabrooke, 'Seeing like an international organization', *New Political Economy*, 17:1 (2012), pp. 1–16.

[73] Pascal Saint-Amans, 'Pascal Saint-Amans finds his stride in Owens's shoes', *International Tax Review*, 23:1 (2012), p. 27.

[74] John Christensen, 'OECD should step aside and let the UN tackle tax havens', *International Tax Review* (2 November 2011), available at: {http://www.internationaltaxreview.com/Article/2928492/OECD-should-step-aside-and-let-UN-tackle-tax-havens.html?ArticleId=2928492} accessed 10 September 2014.

[75] David Roodman, 'The Commitment to Development Index: 2013 Edition' (Washington, Centre for Global Development, 2013), available at: {http://www.cgdev.org/sites/default/files/archive/doc/CDI_2013/Index-technical-description-2013-final.pdf}.

[76] Basel Institute on Governance, 'The Basel AML Index 2013' (Basel: Basel Institute for Governance, 2012), available at: {http://index.baselgovernance.org/sites/default/files/documents/aml-index-project-report-2014.pdf}.

[77] 'The Basel AML Index: Country Risk Ranking' (Basel Institute on Governance, 2014), available at: {http://index.baselgovernance.org/index/Index.html#methodology}; Organisation for Economic Cooperation and Development, 'Bribery and Corruption Awareness Handbook for Tax Examiners and Tax Auditors' (OECD, Paris 2013); 'Nachhaltigkeitsratings: Bankanleihen' (Hannover: imug, 2014), available at: {http://www.imug.de/index.php/nachhaltiges-investment/ratings-bankanleihen.html}; 'Oekom Country Rating: Methodology' (Munich: oekom, 2014), available at: {http://www.oekom-research.com/index_en.php?content=country-methodik} accessed 10 September 2014.

[78] Markus Meinzer, 'White Paper on the FSI 2011', p. 4.

through low tax burdens on business investment and neutrality through a well-structured tax code'.[79]

Furthermore, the FSI has received some support from within the UN system. UNCTAD's 2014 Trade and Development Report relies heavily on the work of TJN and explicitly recognised the FSI's revolutionary content. In outlining the FSI methodology the UNCTAD report notes that with the FSI the 'focus shifts, therefore, from governance issues within countries to the jurisdictions' responsibility for offering financial secrecy at the global level'.[80] Despite mature field opposition and battles over expertise the FSI has passed necessary salience tests for NGO buy-in and there is will in the system to address these complicated tax issues, with some international and private sector organisations offering their support for this revolutionary benchmark.

Conclusion

Benchmarks can push forward reformist and revolutionary agendas, but must face tests of salience, will, and expertise if they are to create normative change. This article presents two ideal types of benchmarking cycles, reformist and revolutionary. The benchmarking cycles differ in steps actors must pass to continue, and especially on the drivers of benchmark creation and the opposing or supporting forces for tests of salience, will, and expertise.

Reformist benchmarking cycles seek adjustment within dominant transnational normative orders. They are pushed by NGOs that invest a great deal of time in carefully selecting issues they believe have salience and political and moral traction. As large and well-resourced NGOs are highly bureaucratised and professionalised organisations, they often lack the expertise to develop benchmarks themselves. Instead they outsource to and consult with experts, or they permit parts of their own organisation to become offshoots following a particular agenda. Once armed with expertise those pushing a reformist agenda also consider stakeholders before making their benchmarks, to fuse expertise with salience. These benchmarks are then taken to political authorities that, with an agenda that strengthens current normative dispositions, can use will to create transnational normative change. As discussed above, we provide two cases of reformist benchmarking cycles, through Transparency International's Corruption Perceptions Index and the Gold Standard's system for carbon credits. Both cases conform to our ideal type reformist benchmarking cycle and both seek to affirm current normative trends.

In contrast to reformist benchmarking cycles, expert-activists drive revolutionary benchmarking cycles. These expert-activists actively contest how we understand how the transnational normative order is composed. They do so by creating benchmarks that directly challenge how we understand how the system works as a whole, as well as how we should analyse the units in the system. Expertise is the key tool for these actors in creating a benchmark, and the obvious weak point for those who wish to oppose it. Expert-activists have to persuade other NGOs to buy-in to their framework

---

[79] Kyle Pomperleau and Andrew Lundeen, 'International Tax Competitiveness Index' (Washington, Tax Foundation, 2014), available at: {http://www.taxfoundation.org/sites/taxfoundation.org/files/docs/TaxFoundation_ITCI_2014.pdf}. In general see Isaac Martin, *Rich People's Movements: Grassroots Campaigns to Untax the One Percent* (Oxford: Oxford University Press, 2013).

[80] United Nations Conference on Trade and Development, 'Trade and Development Report, 2014' (Geneva: UNCTAD, 2014), p. 172.

to past the salience test. They also have to engage in methodological battles with actors in mature fields, who have strong normative, ideational, and material interests for defeating expert-activists in these contests and exposing them as frauds.[81] If expert-activists can get past mature field opposition then they can engage political authorities. This is also a difficult step given that these authorities are also strongly invested in the current normative order on the issue of concern. Opportunities for expert-activists to attempt revolutionary benchmarking are quite common in the NGO community, yet they commonly fail to pass the hurdles designated in our benchmarking cycle due to a lack of NGO buy-in, mature field opposition, and support from political authorities. Sometimes these attempts can garner support and have a political impact. We have provided Tax Justice Network's Financial Secrecy Index as an example of a revolutionary benchmark. As discussed above, it fits with our ideal type, in benchmark creation, NGO buy-in, mature field opposition, and difficulties with political authorities.

The reformist and revolutionary benchmarking cycles include elements of reification, commensuration, focal points, and reactions, the criteria identified by the editors of this Special Issue as common to benchmarking in global governance. As argued, an approach is required that can distinguish what is at stake at points in different cycles, and how actors behave in different ways within the given social structures for benchmark creation and promotion. Benchmarks seek to actively reconfigure how we are governed and also how we seek to govern ourselves. In an era where physical violence between members of the OECD is low, it is through symbolic violence that advanced economies seek to dominate each other, as well as developing economies, as our discussion of 'tax havens' notes. We stress that benchmarks do not represent only the dominant normative order of the time but can contain conforming and contesting views.

Differentiating reformist and revolutionary benchmarks, and locating the tests for salience, will, and expertise, is important for the development of research on activist organisations. It is also important for the theoretical development of scholarship that seeks to understand how fields are composed and how conflict is generated and handled within them. Our work here contributes to Bourdieu-inspired scholarship on how actors are able to transform transnational economic governance,[82] especially where the 'field' is understood as an approach from which to ascertain position-taking within a social environment, rather than as a general theory.[83] Further research could usefully interrogate communication strategies, particularly media strategies to gain salience, which propel revolutionary benchmarks. While reformist benchmarks meet a ready audience, revolutionary benchmarks demand an audience to sit up and pay attention. Here we have provided heuristic tools to identify how different types of organisations must follow distinctive paths to create normative change, as well as how smaller groups can contest dominant forms of symbolic violence through their own expertise and activism.

---

[81] Those perceived as 'knowing well' have a better chance of surviving these contests, see: Leonard Seabrooke and Duncan Wigan, 'Powering ideas through expertise: Professionals in global tax battles', *Journal of European Public Policy*, 23:3 (2016), forthcoming.

[82] The exemplar here is Yves Dezalay and Bryant G. Garth, *The Internationalization of Palace Wars: Lawyers, Economists, and the Contest to Transform Latin American States* (Chicago: University of Chicago Press, 2002).

[83] Mikael Rask Madsen, 'Reflexivity and the construction of the international object: the case of human rights', *International Political Sociology*, 5:3 (2011), pp. 259–75; and Vincent Pouliot, 'The logic of practicality: a theory of practice of security communities', *International Organization*, 62:2 (2008), pp. 257-88.

Benchmarking global supply chains: the power of the 'ethical audit' regime

GENEVIEVE LEBARON AND JANE LISTER*

Abstract. This article critically investigates the growing power and effectiveness of the 'ethical' compliance audit regime. Over the last decade, audits have evolved from a tool for companies to track internal organisational performance into a transnational governing mechanism to measure and strengthen corporate accountability globally and shape corporate responsibility norms. Drawing on original interviews, we assess the effectiveness of supply chain benchmarks and audits in promoting environmental and social improvements in global retail supply chains. Two principal arguments emerge from our analysis. First, that audits can be best understood as a productive form of power, which codifies and legitimates retail corporations' poor social and environmental records, and shapes state approaches to supply chain governance. Second, that growing public and government trust in audit metrics ends up concealing real problems in global supply chains. Retailers are, in fact, auditing only small portions of supply chains, omitting the portions of supply chains where labour and environmental abuse are most likely to take place. Furthermore, the audit regime tends to address labour and environmental issues very unevenly, since 'people' are more difficult to classify and verify through numbers than capital and product quality.

Genevieve LeBaron is Vice-Chancellor's Fellow in Politics at the University of Sheffield. Her research investigates the global growth and governance of forced labour in retail supply chains and the politics of corporate social responsibility. She has published recent journal articles in *Brown Journal of World Affairs*, *Review of International Political Economy*, *New Political Economy*, *International Feminist Journal of Politics*, among other journals. She is co-author, with Peter Dauvergne, of *Protest Inc.: The Corporatization of Activism* (Polity, 2014).

Jane Lister is Senior Research Fellow at the University of British Columbia. Her research examines the environmental policy implications of corporate social responsibility and transnational private governance. She is the author of three books: *Eco-Business: A Big Brand Takeover of Sustainability* (MIT Press, 2013 with Peter Dauvergne); *Corporate Social Responsibility and the State* (UBC Press, 2011); and *Timber* (Polity, 2011 with Peter Dauvergne). Her recent journal articles appear in *Global Policy*, *Global Environmental Change, Millennium*, *Progress in Development Studies*, and *Organization & Environment*.

Introduction

In June 2014, *The Guardian* ran a headline news story revealing the widespread use of 'slavery' and 'human trafficking' by employers in the Thai shrimp

industry.[1] *Guardian* journalists traced the slavery-produced shrimp into the freezers of some of the world's biggest retailers, including Walmart, Tesco, Costco, Aldi, and Morrison's. While the presence of severe labour exploitation in the shrimp industry isn't surprising given retailers' low-cost, high volume business model, what was surprising – and not reported – is that the shrimp had been 'ethically' certified. Indeed, in the wake of *The Guardian's* reports, one of Aldi's key suppliers defensively noted that the shrimp had been certified[2] by GlobalGAP, 'a non-governmental organization that sets voluntary standards for the certification of agricultural products around the globe'.[3] Just as with the 2013 collapse of Bangladesh's Rana Plaza garment factory mere months after it successfully passed a compliance audit,[4] the discovery of severe labour exploitation within a 'certified' supply chain raises crucial questions about the effectiveness and power of global supply chain benchmarks.

Over the last decade, the governance of global supply chains has become increasingly reliant on a 'benchmarking regime'[5] comprised of company codes of supplier conduct, voluntary certifications and standards (for example, Rainforest Alliance Certification), standardised metrics (for example, Higg Index for 'ethical' apparel), and aggregated indexes for comparing corporate environmental and social performance (for example, the Global Reporting Initiative). Through programmes like the Ethical Trade Initiative and Sustainable Apparel Coalition, companies and civil society groups are developing transnational accountability tools like corporate social responsibility (CSR) certifications, which they claim will strengthen environmental and social performance within complex global supply chains. Audit inspections are a cornerstone of these programmes, purportedly allowing multinational retail companies like Walmart and Nike to detect and address problems like environmental degradation or forced labour abuses among their thousands of suppliers. The role and power of audits has grown significantly over the last decade, as audits have evolved from a tool that companies used to track internal organisational performance into a central mechanism of non-state efforts to measure and strengthen corporate accountability globally. Increasingly seen as a way to monitor and improve labour and environmental standards within production, reliance on the audit regime is deepening in the face of inadequate and declining state involvement in global corporate governance.

As we document in this article, over the last decade, global supply chain benchmarks have become increasingly prevalent and influential in shaping corporate responsibility norms, and particularly norms around labour and environmental practices within global production. Retailers claim that their audit-based CSR programmes have advanced both environmental (for example, reduced carbon and waste) and labour (for example, reduced overtime, higher wages) conditions. Yet, as a growing evidence base reveals, audits are ineffective tools for detecting, reporting, or

---

[1] Kate Hodal, Chris Kelly, Felicity Lawrence, Caz Stuart, Thibaut Remy, Irene Baque, Mary Carson, and Maggie O'Kane, 'Globalised Slavery: How Big Supermarkets are Selling Prawns in Supply Chain Fed by Slave Labour', *The Guardian*, available at: {http://www.theguardian.com/global-development/video/2014/jun/10/slavery-supermarket-supply-trail-prawns-video} accessed 10 July 2014.

[2] Intrafish, 'Shrimp were GlobalGAP Certified, Aldi Supplier Says in Wake of Slavery Reports', available at: {www.intrafish.com/news/article1392263.ece} accessed 10 July 2014.

[3] GlobalGAP, 'About Us', available at: {http://www.globalgap.org/uk_en/} accessed 10 July 2014.

[4] Stephanie Clifford and Steven Greenhouse, 'Fast and Flawed Inspections of Factories Abroad', *New York Times*, available at: {http://www.nytimes.com/2013/09/02/business/global/superficial-visits-and-trickery-undermine-foreign-factory-inspections.html?pagewanted=all&_r=0} accessed 10 July 2014.

[5] André Broome and Joel Quirk, 'Governing the world at a distance: the practice of global benchmarking', *Review of International Studies*, 41:5 (2015), pp. 819–41.

correcting environmental and labour problems in supply chains. In fact, they may even be serving to worsen supplier practices as they shape a global business culture of 'check-box compliance' to a narrowing set of quantitative 'key performance measures'.[6] Furthermore, eco and 'ethical' certification may be increasing consumer demand for products that can create further negative social repercussions to workers and communities that fall outside the audit scope.

Although the interdisciplinary literatures on private governance and CSR have analysed shortcomings in audit design and implementation, the global governance implications of growing state, business, and NGO reliance on audits have been widely overlooked. Scholars within management studies and global political economy scholarship, for instance, have investigated and debated whether and under what conditions audits improve environmental and labour practices within supply chains.[7] These debates have evolved along highly technical lines of inquiry whereby audits and associated codes of conduct have tended to be approached as apolitical and neutral instruments. In common with large swathes of liberal scholarship on governance, the underlying assumption has been that due to inadequate state regulatory capacity in many countries in the 'developing world', private governance tools such as audits have emerged as efficient and effective strategies to promote change. Yet, this approach overlooks the role that governments and firms have played in strategically engineering the 'governance gaps' that audits and other forms of CSR have emerged to 'solve', as well as the political nature of private transnational governance more broadly.

This article investigates the deepening role and power of the audit regime, as part of the broader shift towards benchmarking as a form of supply chain governance (see also Tony Porter's article in this Special Issue).[8] Conceptually, it highlights the need to investigate audits and other CSR instruments not as technical or neutral tools, but rather as highly politicised, productive forms of power. Power that is intertwined with the expansion of corporate control and profitability, as well as with 'leadership and expertise intended to sustain and enlarge capitalist market society and its associated

---

[6] See Stephen Davy and Carol Richards, 'Supermarkets and private standards: Unintended consequences of the audit ritual', *Agriculture and Human Values*, 30:2 (2013), pp. 271–81; Hau Lee, Erica Plambeck, and Pamela Yatso, 'Incentivizing sustainability in your Chinese supply chain', *The European Business Review*, May/June (2012); Power, 'Evaluating the audit explosion', *Law & Policy*, 25:3 (2003), pp. 185–202; Richard Locke, *The Promise and Limits of Private Power: Promoting Labor Standards in the Global Economy* (Cambridge: Cambridge University Press, 2013); Ross Taplin, Yafang Zhao, and Alistair Brown, 'Failure of auditors: the lack of compliance for business combinations in China', *Regulation & Governance*, EarlyView DOI:10.111/rego.12011. See also Ole Jacob Sending and Jon Harald Sande Lie, 'The limits of global authority: How the World Bank benchmarks economies in Ethiopia and Malawi', *Review of International Studies*, 41:4 (2015), pp. 993–1010.

[7] Gary Gereffi, 'Can Global Brands Create Just Supply Chains?', *Boston Review,* available at: {http://www.bostonreview.net/forum/can-global-brands-create-just-supply-chains/host-countries-can-act} accessed 10 July 2013; Richard Locke, *The Promise and Limits of Private Power*; Michael Power, *The Audit Society: Rituals of Verification* (Oxford: Oxford University Press, 1997); Power, 'Evaluating the audit explosion', pp. 185–202; Michael Toffel, Jodi L. Short, and Melissa Ouellet, *Reinforcing Regulatory Regimes: How States, Civil Society, and Codes of Conduct Promote Adherence to Global Labor Standards*, Harvard Business School Technology and Operations Management Unit Working Paper 13-045, available at: {http://www.hks.harvard.edu/m-rcbg/CSRI/publications/workingpaper_65_toffel_short_ouellet.pdf} accessed 3 April 2014.

[8] As distinct from Tony Porter who examines the incremental capabilities of benchmarking as a potentially effective 'accelerator' or information 'relay' management mechanism in improving business practice within collaborative governance networks, we focus on the broader more fundamental procedural and normative global regulatory limits of benchmarking (as the tenet of the audit regime) in blocking more transformative social and environmental change. See Tony Porter, 'Global benchmarking networks: the cases of disaster risk reduction and supply chains', *Review of International Studies*, 41:5 (2015), pp. 865–86.

principles of governance'.[9] We argue that far from helping firms, governments, and consumers to 'monitor' and address social and environmental problems, the audit regime is serving to stabilise, legitimise, and conceal endemic problems within supply chains. More specifically, we argue that many audit systems are designed to be ineffective at detecting and communicating the environmental and labour abuses fundamentally associated with the global retail business model, and that powerful business interests leverage for their own gains highly strategic control over *how* and *when* audits are conducted, and *what* is evaluated. In developing this argument, we synthesise across recent large-scale studies of audit ineffectiveness, as well as draw on original interviews with ethical auditors, retail buyers, consumer goods suppliers, NGO representatives, and CSR managers and experts from the United Kingdom (UK), the United States (US), and China, as well as factory visits in and around the Pearl River Delta region of China.[10]

The article unfolds in four parts. In the first section, we document the rise of the audit regime and argue that it can be best understood as a structural and productive form of power[11] that codifies and legitimates retail corporations' poor social and environmental records and shapes state approaches to supply chain governance. In the second section, we argue that growing public and government trust in the metrics generated by audits ends up concealing real problems in global supply chains. Retailers are, in fact, auditing only small portions of supply chains, and tend to omit the portions of supply chains where labour and environmental abuse are most likely to take place. In section three, we assess variation in audit effectiveness regarding environmental and labour practices. We argue that the audit regime tends to address labour and environmental issues unevenly, since 'people' are more difficult to classify and verify through numbers than capital and product quality. We conclude by considering how, as NGO involvement is expanding political traction of social and environmental benchmarks verified through audits, the audit regime ultimately disguises a normative, market-based policy agenda in seemingly objective tools and metrics.

Audits as a productive form of power

Corporate use of compliance audits as internal tools to examine and measure organisational non-financial performance stretches back to the 1980s and 1990s, as

---

[9] Stephen Gill (ed.), *Global Crises and the Crisis of Global Leadership* (Cambridge: Cambridge University Press, 2012), p. 1.

[10] Secondary data collection for this article involved a desk-based review of the audit industry, with a focus on audit programmes, protocols and retail companies' codes of conduct for suppliers. Primary data collection involved semi-structured elite interviews. A total of 23 in-person interviews and 2 telephone interviews were conducted in China, North America, and the United Kingdom with retail buyers, auditors, consumer goods suppliers, factory managers, NGO representatives, trade unionists, and CSR managers and experts. A site visit to Guangdong province, China, took place in April 2012 and interviews and factory visits were conducted at and around the China Import and Export Fair ('The Canton Fair'). The UK interviews were conducted in March 2013, in and around London. The North American interviews were conducted in Vancouver (Canada) and Seattle (US) between May 2012 and July 2013. The interview and factory visit data presented here is intended to be illustrative, rather than comprehensive.

[11] We are indebted to Michael Barnett and Raymond Duvall's conception of productive power as 'the constitution of all social subjects with various social powers through systems of knowledge and discursive practices of broad and general social scope'. See Michael Barnett and Raymond Duvall, 'Power in international politics', *International Organization*, 59 (Winter 2005), pp. 55–7. We are indebted to neo-Gramscian global political economy scholarship for our conception of the structural power of business and audits. See, for example, Stephen Gill (ed.), *Power and Resistance in the New World Order* (New York: Palgrave, 2003); Stephen Gill (ed.), *Gramsci, Historical Materialism, and International Relations* (Cambridge: Cambridge University Press, 1993).

governments outsourced and encouraged industry self-regulation.[12] But the adoption of audits as a global corporate regulatory mechanism is much more recent. Alongside other benchmarking tools, the rise of audits has taken place in the context of a 'decentering' of regulation and enforcement of labour and environmental standards away from states.[13] These processes have occurred as part of broader transformations in the global economy, including the rise of neoliberal forms of globalisation.[14] As has been widely documented in the global governance literature, such transformations have entailed dramatic changes in approaches to economic and corporate regulation as states have redefined their relationships with market actors and especially with transnational corporations. As governments have conferred 'privileged rights and citizenship and representation on corporate capital',[15] and opened the door to the heightened involvement of private actors in economic governance processes, corporations have increasingly set and enforced their own standards and rules. Today, in the context of the privatisation and marketisation of governance, an increasingly 'significant degree of global order is provided by individual firms that agree to cooperate, either formally or informally, in establishing an international framework for their economic activity'.[16]

As transnational corporations have become increasingly entrusted to govern themselves, and report on their efforts to government and the public,[17] there has been a persistent decline of state-based monitoring of production processes in many countries. For instance, as the International Labour Organisation has documented, there has been a steep downturn in the number of labour inspections conducted in countries in both the global North and South, and in some instances, the outright elimination of the labour inspectorate.[18] Similarly, scholars in global environmental

---

[12] David Osborne and Ted Gaebler, *Reinventing Government: How the Entrepreneurial Spirit is Transforming the Public Sector* (New York: Plume, 1993); Power, *The Audit Society*.

[13] cf. Julia Black, 'Decentering regulation: Understanding the role of regulation and self regulation in a "post regulatory" world', *Current Legal Problems*, 54 (2001), pp. 103–46; See, for example, Tim Büthe and Walter Mattli, *The New Global Rulers: The Privatization of Regulation in the World Economy* (Princeton: Princeton University Press, 2011); Claire Cutler, 'Private transnational governance and the crisis of global leadership', in Stephen Gill (ed.), *Global Crises and the Crisis of Global Leadership* (Cambridge: Cambridge University Press, 2012), pp. 56–70.

[14] For an overview of neoliberal globalisation's discursive and material characteristics, see Jamie Peck, *Constructions of Neoliberal Reason* (Oxford: Oxford University Press, 2012).

[15] Stephen Gill, 'Globalisation, market civilisation, and disciplinary neoliberalism', *Millennium Journal of International Studies*, 24:3 (1995), pp. 413. See also Claire Cutler, Virginia Haufler, and Tony Porter, *Private Authority and International Affairs* (Albany: State University of New York Press, 1999). See also James Harrison and Sharifah Sekalala, 'Benchmarking human rights at the United Nations: Self-reporting by states and corporations', *Review of International Studies*, 41:5 (2015), pp. 925–45.

[16] Cutler, Haufler, and Porter, *Private Authority and International Affairs*, pp. 3–4.

[17] The recent wave of audit-based approaches to 'modern slavery' and forced labour – developed by multi-stakeholder coalitions of NGOs and retail and brand corporations – are illustrative. In 2012, California passed the 'Transparency in Supply Chains' (TISC) Act, which requires large companies to report on what they are doing to verify their global supply chains against forced labour, trafficking, and slavery. The TISC Act institutionalises audit-based approaches to dealing with severe labour exploitation, requiring companies to report on their voluntary efforts to detect and address severe exploitation, rather than mandating specific standards or benchmarks and requiring companies to demonstrate compliance. Similarly, in the United Kingdom, the draft Modern Slavery Bill currently before Parliament deepens the 'light-touch' approach to business regulation in that country, and includes only a cursory treatment of supply chains. In a recent press release entitled, 'Government Asks Retailers to Lead the Way on Transparent Supply Chains', the government announces plans to work with retailers on a 'best practices' report, centered around 'ethical audit programmes, outlining some of the main certification schemes and collaborative initiatives'. See UK Department for Business, Innovation & Skills, 'Government Asks Retailers to Lead the Way on Transparent Supply Chains', available at: {https://www.gov.uk/government/news/government-asks-retailers-to-lead-the-way-on-transparent-supply-chains} accessed 4 August 2014.

[18] Renato Bignami, Giuseppe Casale, and Mario Fasani, *Labour Inspection and Employment Relationship* (Geneva: International Labour Organization, 2013).

politics have documented persistent declines in state-based environmental compliance efforts.[19] Within this context, as we analyse in the remainder of this section, audits have emerged as a key tool of corporate supply chain governance. We argue that although the corporate, NGO, and state motivations for institutionalising the audit regime vary, the adoption of audits nevertheless reflects an alignment of these actors' interests and ideologies in market-based global governance. We argue further that in codifying and legitimating corporations' poor environmental and labour records, the audit regime is playing a productive role in shaping public and policymakers' perceptions of corporate practices, as well as corporate power globally.

Adopting audits

The shift towards self-regulation of corporations has intersected with the growing willingness of certain advocacy organisations to work alongside transnational corporations towards environmental and social goals.[20] The reasons for the proliferation and deepening of partnerships between non-governmental organisations (NGOs) and firms are complex. For some organisations, frustration with decades of slow progress and failure to achieve significant 'scale' in the outcomes of their social and environmental advocacy efforts has prompted them to shift their strategies towards market-based approaches. For others, in states like Canada and the US, growing amounts of government funding for NGO activities has come to depend on corporate involvement, making corporate collaboration as a necessity for continued operation. In short, although NGOs' interests in partnering towards initiatives – and supply chain benchmarking more broadly – vary across organisation cause, size, region, at a general level it is clear that many organisations have either chosen or been forced by governments to join forces with corporations, and have come to engage in activities that legitimate rather than challenge highly unequal and unsustainable patterns of global production. These shifting relationships between states, civil society, and corporations have been driven by broad socioeconomic changes that are, as Peter Dauvergne and Genevieve LeBaron describe, 'reconfiguring power and resistance globally, as firms engage social forces through corporate social responsibility, as governments cut social services and devolve authority to companies, as consumerism spreads, and as states suppress public dissent'.[21]

In the context of these reconfigurations, large numbers of NGOs are now attempting to modify corporate practices through 'private governance' initiatives that depend heavily on audits.[22] Projecting a positive image to the public of taking direct action to come up with solutions with the biggest companies contributing to the

---

[19] For an overview, see Peter Newell, *Globalization and the Environment: Capitalism, Ecology, and Power* (Cambridge, UK: Polity, 2012).

[20] Peter Dauvergne and Jane Lister, *Eco-Business: A Big-Brand Take-over of Sustainability* (Cambridge, MA: The MIT Press, 2013); Peter Dauvergne and Genevieve LeBaron, *Protest Inc.: the Corporatization of Activism* (Cambridge: Polity, 2014); Peter Newell, 'Managing multinationals: the governance of investment for the environment', *Journal of International Development*, 13:7 (2001), pp. 907–19; Philipp Pattberg, 'The institutionalization of private governance: How business and nonprofit organizations agree on transnational rules', *Governance*, 18:4 (2005), pp. 589–610.

[21] Dauvergne and LeBaron, *Protest Inc.*, p. 2.

[22] As Robert Falkner explains, '"private governance" emerges at the global level where the interactions among private actors or between private actors on the one hand and civil society and state actors on the other, give rise to institutional arrangements that structure and direct actors' behavior in an issue-specific area.' See Robert Falkner, 'Private environmental governance and International Relations: Exploring the links', *Global Environmental Politics*, 3:2 (2003), p. 72.

biggest problems, in the past five years, NGOs have increasingly partnered with firms to develop or implement voluntary CSR programmes: Greenpeace has worked with Coca Cola to reduce greenhouse gas emissions; Conservation International has worked with Walmart to track and mitigate illegal sourcing for its jewellery products; Oxfam has partnered with Unilever to integrate smallholder farmers 'fairly' into transnational supply chains. Explaining why they are working so closely with Walmart, the head of Conservation International explains, 'Given the millions of items carried by its thousands of stores, possibilities virtually are endless for the company to create extraordinary impact.'[23]

Many NGO-business programmes rely on benchmarking practices, including audit inspections – rather than state labour inspectors or environmental agencies, or codified legal agreements – to test and ensure compliance and effectiveness to certifications (for example, Forest Stewardship Council), standards (for example, Fairtrade), and retailers' own sustainability commitments (for example, Sainsburys' 20 by 20 Sustainability Plan).[24]

Simply put, compliance auditing involves independently verifying supplier performance to corporate codes of conduct, often through the use of 'independent' but for-profit firms like SGS or UL, hired by the brand buyer. As NGOs have embraced private governance as a means to 'scale up' and strengthen corporate accountability – and thus, broadly improve labour conditions and preserve the environment – they have also adopted audits as a global corporate regulatory mechanism.

The standards and certification schemes proliferating among multi-stakeholder coalitions are audit-based. For instance, the Extractive Industries Transparency Initiative coalition – which brings executives from Chevron and Shell together with NGOs like Global Witness, Oxfam, and Transparency International – has developed the EITI standard validated by audit firm Deloitte, among others.[25] Similarly, through the Fair Labor Association (FLA), organisations like the Global Fairness Initiative, Maquila Solidarity Network, and Human Rights First[26] have worked with big brand partners including Nestlé, Nike, Adidas, Apple, and H&M to develop the FLA Workplace Code of Conduct.[27] Adherence to the FLA's Code is enforced through independent social audit firms, including Verité and Impactt.[28]

Some NGOs have also come to use supply chain benchmarks to 'raise awareness' about their causes. They also now depend on the metrics generated through these programmes for outreach and fundraising efforts, which further explains NGO interest and support for the audit regime. The Greenpeace campaign to reduce the

[23] Peter Seligmann, 'Conservation international's CEO: Why Walmart gives me hope', *GreenBiz* (9 May 2014), available at: {http://www.greenbiz.com/blog/2014/05/09/retail-giant-walmart-asks-suppliers-shrink-environmental-impact} accessed 8 April 2015.

[24] Given corporate, NGO, and state involvement in the audit regime, within André Broome and Joel Quirk's typology of global benchmarking practices (this Special Issue), audits can be situated at the cusp of type III (private market governance) and type IV (transnational advocacy). They can involve both profit-based and civil society organisations as monitoring agents, but seek to actuate transnational market governance.

[25] Extractive Industries Transparency Initiative, 'EITI Validators', available at: {http://eiti.org/validation/validators} accessed 4 August 2014.

[26] Fair Labor Organization, 'FLA Accredited Monitoring Organizations', available at: {http://www.fair-labor.org/transparency/fla-accredited-monitoring-organizations} accessed 4 August 2014.

[27] Nestle, 'Annual Report 2011', available at:{http://static.globalreporting.org/report-pdfs/2012/537fc4b7fef52936fc433f1f6a77a496.pdf} accessed 4 April 2014.

[28] Fair Labor Organization, 'FLA Accredited Monitoring Organizations', available at: {http://www.fair-labor.org/transparency/fla-accredited-monitoring-organizations} accessed 4 August 2014.

environmental impact of electronics for example, is centered on producing a widely distributed *Guide to Greener Electronics* – a ranking of the 'environmental leadership efforts' of brand electronics companies based on a score out of 10 on 12 criteria within 3 (shifting) categories of environmental impact (energy, products, operations). Greenpeace conducts their benchmark evaluation using the public reports and information that the brand companies produce. This has lead to criticism (particularly from lower ranked companies such as Apple) that the campaign evaluates companies for what they say rather than what they do. However, despite concerns about its accuracy or effectiveness, their metrics-based report has given Greenpeace much media attention and a positive impression with the general public (with the reporting of specific company performance figures), that Greenpeace is 'cracking the whip' to monitor and hold corporations to account.[29] Meanwhile, the exercise of metrics production has distracted attention by all sides away from the major underlying environmental issue – planned obsolescence in electronics product design.

Given growing NGO buy-in, it is perhaps not surprising that at the global level, too, benchmark audit-based initiatives are accelerating. The United Nation's Global Compact is premised upon the use of audit-based initiatives to 'promote transparency and disclosure as a means of driving performance'.[30] Functioning as a 'learning network', the Global Compact encourages 'desirable behavior in corporations through dialogue with different stakeholders and sharing of information'.[31] The associated Global Reporting Initiative (GRI) has developed and disseminates 'Sustainability Reporting Guidelines', which companies can implement among their suppliers, as well as assurance criteria for external audit verification to determine whether the environmental, social, and governance (ESG) data have been reported accurately.[32] Financial institutions and states increasingly rely on these as proxy indicators of corporate responsibility (particularly in offshore operations).[33]

Reflective of broader trends in the privatisation of global corporate governance, states have helped to institutionalise voluntary, CSR-based forms of corporate governance. Some of the environmental and social standards that retailers achieve through audits – such as ISO 14000 – have actually been adopted by states and international organisations as official standards, thus gaining 'in strength and legitimacy'. As Robert Falkner describes, 'governments are expected to incorporate them into their procurement and international bodies such as the WTO have recognized the voluntary ISO standards as international standards under the WTO system and as being consistent with the Technical Barriers to Trade Agreement.'[34] Certain states have also facilitated partnerships and coalitions, as the United States did with the Climate Action Partnership (USCAP). This coalition brings

[29] See 'Greenpeace Electronics Guide Cracks the Whip', BlueChannel 24, available at: {http://www.blue-channel24.com/?p=7749} accessed 8 April 2015.

[30] United Nations Global Compact, 'Analyzing Progress', available at: {http://www.unglobalcompact.org/COP/analyzing_progress.html} accessed 5 November 2013.

[31] Susanne Soederberg, 'Taming corporations or buttressing market-led development? A critical assessment of the global compact', *Globalizations*, 4:4 (2007), p. 503. Significantly, like other global transparency programmes that encourage and facilitate corporate voluntary reporting, the Global Compact allows companies to set their own measures and report on progress towards these, rather than setting universal benchmarks. Performance comparisons between companies are therefore difficult.

[32] Global Reporting Initiative, 'The External Assurance of Sustainability Reporting' (2013), available at: {https://www.globalreporting.org/resourcelibrary/GRI-Assurance.pdf} accessed 8 April 2015.

[33] David Vogel, 'The private regulation of global corporate conduct: Achievements and limitations', *Business & Society*, 49:1 (2010), pp. 68–87.

[34] Falkner, 'Private environmental governance', p. 77.

environmental organisations such as the Environmental Defense Fund, The Nature Conservancy, and the World Resources Institute, together with Shell, NRG Energy, as well PepsiCo, DuPont, The Dow Chemical Company, Rio Tinto, and General Electric.[35] As networks and institutions become shaped around audit-based accountability systems, the prevalence of 'ethical audits' can be expected to increase not diminish.

The power of audits in transnational corporate governance

The growing adoption of audits and associated benchmarking regimes as a corporate governance mechanism is primarily an effort to mediate the labour and environmental risks inherent in the retail-driven, low-cost, high volume, model of distant global production that currently reigns in the global economy. As big brand firms and retailers have grown rapidly in recent years – with Walmart's total global sales nearing $500 billion – they have pioneered a business model in which they coordinate the production of goods through loose and arm's length relationships with tens of thousands of independent suppliers, but don't actually produce goods themselves. The acceleration of ever-more complex forms of subcontracting has meant that global retail supply chains are increasingly long and complex with many 'tiers' of outsourced production.[36] Walmart, for instance, now sources from over 100,000 global suppliers, who source from a diffuse and elaborate global production network.[37] Simply put, the retail model of global production has developed to maximise lead firms' flexibility and continuously reduce costs and liability associated with production. As retailers have grown in size and market power, they have exerted downward pressure on contract length, turnaround times, and margins in many industries, 'squeezing' their suppliers. This business model and power dynamics introduce endemic labour and environmental risk into global supply chains, as supplier firms struggle to fulfill orders on time and at a low-cost, creating incentives to cut corners.[38]

Retailers and big brands initially introduced auditing as an internal tool to monitor and manage risks within their global operations, examining and measuring organisational non-financial performance, as well as key suppliers. But as NGO and journalist exposés of rampant child and 'sweatshop' labour and environmental malpractice in overseas production sites yielded calls for greater corporate transparency and accountability beginning in the late 1990s and throughout the early 2000s, brands began to adopt audits not as an heuristic device, but as tools to mediate and resolve the tensions and contradictions inherent in outsourced, low-cost, quick turnaround production. In the wake of persistent non-compliance issues, brands began hiring independent (but often for-profit) audit firms to monitor supplier factories, such as Apple most recently did with its Tier 1 supplier Foxconn in the wake of a wave of worker suicides and reports of forced overtime.[39] Given the need to

[35] United States Climate Action Partnership, available at: {http://www.us-cap.org/} accessed 4 August 2014.

[36] William Milberg and Deborah Winkler, *Outsourcing Economics: Global Value Chains in Capitalist Development* (Cambridge: Cambridge University Press, 2013); see also: Layna Mosley, *Labor Rights and Multinational Production* (New York: Cambridge University Press, 2011).

[37] Walmart, 'Walmart Announces Sustainable Product Index', available at: {http://news.walmart.com/news-archive/2009/07/16/walmart-announces-sustainable-product-index} accessed 4 August 2014.

[38] See, for example, Milberg and Winkler, *Outsourcing Economics*.

[39] See, for example, Stanley James and Adam Satariano, 'Apple Opens Suppliers' Doors to Labor Group After Foxconn Worker Suicides' (13 January 2014), available at: {http://www.bloomberg.com/news/2012-01-13/apple-opens-suppliers-doors-to-labor-group-after-foxconn-worker-suicides.html} accessed 1

govern global corporate practices in a world lacking global regulation, and in the face of declining state-based corporate monitoring, the auditing regime has stepped into the lurch, with experts and the organisations driving adoption (as described above) claiming that it promotes supply chain transparency, accountability, and improvements.

The difficulty, however, is that audits are an inadequate mechanism for ensuring global corporate accountability. While, as we document, this dynamic relates to weaknesses in the audit regime's design and functioning, it is important to note here that audit ineffectiveness is also rooted in and reflects the tensions and limits of benchmarking as a form of supply chain governance 'at a distance'. As critical scholars of the audit regime such as Dara O'Rourke and Richard Locke argue[40] and as our interviewees confirm, auditing produces standardised metrics, measurements, and rankings that create the appearance of independent supply chain monitoring; yet, the information produced through and derived from audits is partial, highly political, and fundamentally shaped by the retail audit client.

As one auditor described, most audit firms have no investigative powers, so have limited capacity to verify that the information presented to them is accurate: 'you have no powers of search so you cannot open a locked drawer, you cannot check to see how much tax is being paid or has actually been paid, you can look at a record that says something but you wouldn't be able to go and find out whether it's actually true'.[41] As another auditor argued, the audit regime is not actually designed to drive improvements in supplier practices: 'an audit is a diagnostic tool; it doesn't fix things. It doesn't matter how many times we audit a factory, it doesn't mean they're going to improve.'[42] The former Director of CSR for a major US outdoor gear retail company emphasised the limits very clearly: 'Within the social compliance world, it is now standard operating understanding that audits don't work to achieve change within organizations. That's widely discussed from folks at the Gap to Patagonia to you name it. In the NGO community, many social activists would also say that the results indicate that audits don't work in terms of actually effecting real change on the ground.'[43]

Indeed, as a number of recent studies have documented and as companies have acknowledged, the capacity of the audit regime to detect, report, and correct non-compliance is limited.[44] Yet, in spite of mounting evidence of its ineffectiveness, the audit regime is extending and gaining legitimacy as a global regulatory mechanism, as

---

September 2014; Susan Adams, 'Apple's New Foxconn Embarrassment', *Forbes* (9 December 2012), available at: {http://www.forbes.com/sites/susanadams/2012/09/12/apples-new-foxconn-embarrassment/} accessed 1 September 2014.

[40] Dara O'Rourke, 'Monitoring the monitors: a critique of corporate third-party Labour monitoring', in Rhys Jenkins, Ruth Pearson, and Gill Seyfang (eds), *Corporate Responsibility and Labour Rights: Codes of Conduct in the Global Economy* (London: Earthscan, 2002), pp. 196–207; Locke, *The Promise and Limits of Private Power.*

[41] Personal communication with auditor, London, 6 March 2013.

[42] Personal communication with audit firm Director of Sustainability, London, 13 March 2013.

[43] Personal communication with a former Director of CSR, US retail company, 12 July 2013, Seattle.

[44] Richard Locke at MIT's 4-year analysis of Nike's audit programme (2001–5) found that workplace conditions in almost 80 per cent of suppliers remained the same or worsened over time. Nike states in their 2012 Sustainable Business Report, '… we have learned that monitoring does not bring about sustainable change. Often, it only reinforces a pattern of hiding problems.' Similarly, HP concludes in their 2013 CSR report, 'although audits can be an excellent measurement tool, they only provide a snapshot of performance and do not lead to lasting performance improvements on their own'. See also Jean Allain, Andrew Crane, Genevieve LeBaron, and Laya Behbahani, *Forced Labour's Business Models and Supply Chains* (York: Joseph Rowntree Foundation, 2013).

NGOs and governments endorse and delegate authority to audit mechanisms as part of the broader consensus around the effectiveness and efficiency of market-based approaches to global governance. Through increasingly sophisticated metrics, and the inclusion of key expert communities, the audit regime is creating the appearance of global supply chain governance and 'continuous improvement' when it in fact serves to legitimate business as usual. As one of our informants explained, 'there is a whole industry of ethical auditors out there now who will find nothing if you pay them to go and find nothing'.[45] In another informant's words, many auditors 'are not trying to find things out, they are trying to prove that something is not there'.[46]

As auditors are bound by rigid confidentiality clauses, their retail clients exercise full discretion over what audit information is ultimately reported. Information is typically shared with the supplier firm but is rarely made available to governments or the public. This means that issues like forced labour, human trafficking, or pollution are not reported to regulators or consumers, and are rarely resolved. Auditors typically offer expert advice to aid the audited factory in preparing an action plan to address the non-compliance findings.[47] However, the auditor has no influence over the company's eventual business decisions and there is no external accountability for the action plans. The audit information that is publicly circulated is typically in the form of general metrics and aggregated analysis that effectively provide neutralised accounts of corporate practices to governments, investors, and consumers. These reports produced by retailers, experts, and multi-stakeholder coalitions consequently serve to codify and legitimate rather than expose and challenge endemic, poor environmental and social problems that lie at the heart of the retail business model.

The independent power of the audit regime

Before turning to explore the inadequacies of the audit regime in greater institutional detail, it is important to note that although insiders recognise the audit system is grossly inadequate, the knowledge generated through the audit system plays a productive role in shaping public and policymakers' perceptions of corporate practices. Corporations are increasingly aware of this, and are investing in devising, producing, and publicising new metrics not only as legitimation techniques to prevent more stringent restrictions on their activities, but also as a tool for sales and competitive advantage.

Revealingly, in 2014, the global industry organisation Sustainable Brands – which features Nestlé, Mars, Coca Cola, Target, Unilever, and Disney as members – convened a conference on 'New Metrics'. The conference promised to convene business leaders to 'define the business metrics of the future … [to] find new approaches to value risk and impact that lead to successful business results'. The '10 Reasons to Attend New Metrics 14' advertised to potential attendees included the following: 'There are growing opportunities to quantify brand value added through innovation for sustainability; Consumers are starting to bypass or boycott clearly unsustainable products and services; Leading methodologies for quantifying both environmental and social impacts are getting sophisticated; Your sustainability performance will be ranked by external stakeholders anyways, so it's a good idea

---

[45] Personal communication with labour NGO representative, London, March 2013.
[46] Personal communication with auditor, London, 6 March 2013.
[47] Salo Coslovsky and Richard Locke, 'Parallel paths to enforcement: Private compliance, public regulation, and labour standards in the Brazilian Sugar Sector', *Politics & Society*, 41:4 (2013), pp. 497–526.

to lead; You could be influencing policy makers and politicians more effectively, armed with new impact assessments.'[48] In short, this discourse reflects a growing recognition by corporations of the utility of supply chain benchmarks – not to change their business practice so much as to transform *understanding* of their practices, and in so doing, generate sales, influence regulatory outcomes, and capture greater business gains.

Corporate interest in generating new metrics is grounded in a business desire to control performance information so as to shape reputational perceptions and expectations of company value. Metrics enable industry to strategically present normative, partial, and fragmented information about their practices in a way that is viewed as scientific and objective and hence, trustworthy and reliable. As legal scholar Galit Sarfaty explains, 'the use of quantitative indicators can be fraught with problems, which are often overlooked due to the authoritative quality of numbers'.[49]

These insights point to the need to understand the audit regime as a productive form of power, which is 'changing understandings, meanings, norms, customs, and social identities that make possible, limit, and are drawn on for action'.[50] The audit regime is not only bolstering corporate power and influence, but the generated knowledge and associated metrics are also increasing corporations' 'public and private authority in the international realm by enabling indirect forms of power that operate through abstract knowledge practice'.[51] Ultimately, although the audit-based data reported by retailers obscures more than it reveals about environmental and social problems within supply chains, the sheer volume of information and metrics produced through audits, and the scientific authority constructed through the involvement of 'independent' experts, 'represent and reify particular normative standards by rendering ideas about appropriate modes of conduct into observable data'.[52] In the context of neoliberal globalisation, these discursive trends are paralleled by and serve to deepen the power and profitability of firms.

Compliance audits and the concealment of ethical problems

Despite their ineffectiveness, compliance audits continue to gain legitimacy as a mechanism of global corporate accountability due to their perceived neutrality as anchored in supposedly objective metrics. However, as this section shows, audits are far from impartial. Rather, the problems with audit effectiveness in this respect manifest around the politicisation of audit design and scope by powerful business interests. This includes highly strategic corporate control over both *how* audits are conducted as well as *what* they are evaluating. Given the evidence presented in this section, we argue that audits are serving to reinforce and legitimise rather than challenge the inequities in power relations within global value chains that are at the root of continually occurring environmental and labour problems. While compliance auditing has increased over the past decade, it has done little to transform corporate practices or improve overall environmental conditions. Focused on limited, incremental adjustments to operations, the compliance audit regime is ultimately

---

[48] Personal communication with Sustainable Brands, 2 September 2014.
[49] Galit Sarfaty, 'Regulating through numbers: a case of corporate sustainability reporting', *Virginia Journal of International Law*, 53 (2011), pp. 580–1.
[50] See Barnett and Duvall, 'Power in international politics', p. 56.
[51] André Broome and Joel Quirk, 'Governing the world at a distance'.
[52] Ibid.

allowing companies to stave off efforts to impose more stringent public regulation and truly accountable restrictions on production.

Procedural politics of audit design

Every multinational brand retail company has their own, unique 'ethical compliance' audit programme to manage their business risks and increase their control over their suppliers. Although many retailers share the same suppliers and source products from the same offshore factories, their audit processes vary. The main procedural differences include: the timing of the audit, the auditors selected, and the communication of results. Decisions around each of these factors are highly strategic as they play a significant role in determining the audit findings which companies then shape and communicate for reputation management and legitimacy (social license) to operate and grow.

Strategic scheduling

Decisions on audit scheduling have significant bearing on audit results. The time of year, frequency of audits, and whether it is an announced or surprise audit will affect what is revealed or not. An audit conducted during peak production cycle before Christmas, for example is much more likely to uncover 'non-conformance incidents' such as illegal workers and unacceptable working conditions than during slower off-peak periods. A more frequent audit than one every few years is also much more likely to uncover problems and yield greater understanding to trace and address the source of the factory conditions. And finally, whether the audit is a 'surprise' versus scheduled in advance will influence the extent to which the factory can potentially obfuscate problems (for example, falsify records, shift labour, move production to a shadow company, etc.). Scheduling elements are not standardised across company audits. Rather, companies 'play' with varying these audit design elements depending on their interests to expose a more or less accurate picture of supplier factory conditions.

Controlled audit delegation

The independence and expertise of the audit team also bears on the results of the audit and companies select their auditors for strategic reasons depending on what they want or do not want to find out about their suppliers. This includes varying whether the auditors are internal to the company or constitute an independent third party; and whether they are local experts or professionally designated auditors brought in from outside.

Because of the proliferation of voluntary compliance standards and certification schemes over the past two decades, there are now many accredited audit firms to select from. They conduct audits to company codes as well as a wide range of social and environmental certification standards. In agriculture alone, there are 'more than 300 accredited certification bodies around the world providing inspection and certification services to organic farmers and producers' (50 per cent do other audits as well).[53] The large professional accounting firms such as PwC, Deloitte, KPMG, and Ernst & Young also all have accredited groups that conduct sustainability audits.

---

[53] International Organic Accreditation Service (IOAS), 'Certification Body Database', available at: {http://www.ioas.org/certification_body/} accessed 4 August 2014.

For assessing compliance, companies choose whether to contract out to an independent auditor or whether to utilise internal audit teams. An internal team enables greater control over the audit process but may also compromise perceived objectivity, potentially threatening some of the reputational capital companies aim to gain from having an audit programme. A third party auditor is generally deemed more neutral and hence, legitimate. Yet, even with third party auditors, the audit process in not impartial. Walmart, for example, applies its own criteria for selecting a list of whom it deems 'acceptable' auditors from which their suppliers may only choose.

Politics also come into play in the development of audit protocols (that define the metrics and benchmarks for compliance). The interpretation of protocols is often heavily shaped by the auditor/client relationship. Whether a professional global accounting firm or local audit business, the auditor is never free of the potential for conflict of interest and bias from wanting to modify audit results in order to obtain payment and retain the business contract.[54] This may mean reporting 'corrective action' audit results in a more positive or negative light (that is, minor versus major non-conformance) depending on the interests of the client (for example, the retail buyer, the supplier, or the certification body). Certification auditors may, in particular be influenced towards leniency by the need to gain or maintain corporate participants in their programmes (to gain scale and legitimacy of their standards) rather than outright 'disqualify' companies that fail to meet the standards. The inconsistency of Forest Stewardship Council audits and continuing challenges to many of the certifications awarded provide a good example of the varying auditor expertise and the politics of auditor conflict of interest.[55]

Constructed transparency

Although lauded as a means to improve corporate transparency and accountability, as mentioned, audit results reports are in fact private information – held confidentially between the client and auditor. Findings that are communicated publically are strategically aggregated so as to conceal the location, type and size of the factory, as well as the scope and criteria of the audits and precise description of the non-compliance problem encountered. Even the names of suppliers are typically withheld. Nike consolidates factory audit data into a single 'sourcing and manufacturing sustainability index' to measure the extent to which a contracted factory is 'lean, green, equitable and empowered'.[56] Many companies provide audit results as percentages of high, medium, or low compliance by geographic region with comparisons in percentage totals between years. The factories audited, however, differ from year to year so the numbers reported do not actually convey any sort of useful or consistent information as to whether conditions at certain factories are improving or worsening and if so, in what way. The reported audit metrics are carefully shaped to give the illusion of progress.

[54] See the recent study of 44,383 social audits in 47 countries by Michael Toffel and colleagues at the Harvard Business School that finds 'evidence suggesting that violations recorded in audits might indeed be influenced by financial conflicts of interest and by auditor competence'. (Michael Toffel, Jodi Short, and Melissa Ouellet, 'Codes in context: How states, markets, and civil society shape adherence to global labor standards', *Regulation & Governance*, doi:10.1111/rego.12076 (2015), p. 16.

[55] Jane Lister, *Corporate Social Responsibility and the State* (Vancouver: UBC Press, 2011).

[56] Nike, 'Sustainable Business Report' (2013), available at: {http://www.nikeresponsibility.com/report/files/report/NIKE_SUSTAINABLE_BUSINESS_REPORT__FY10-11_FINAL.pdf} accessed 3 April 2014.

Material limits in audit scope

It is not just the discretion companies wield with *how* compliance audits are conducted that conceal problems but also perhaps even more significantly, the accepted power companies have to shape *what* the auditors evaluate.

Facilities focus

The scope of compliance audits is not standardised. Some consist of a minimal offsite desk review of documentation and records; others a telephone interview with management; and others (most frequently), an onsite factory investigation. The greater the number of days onsite and the larger the number of interviews with management, employees and outside stakeholders (for example, community members, subcontractors, local government), the greater the costs and the more comprehensive the audit findings. Companies closely control the scope of the audit based on their resources and the extent that they want to know the problems going on.

Increasingly, retailers are designing their audit focus around a risk scale. HP's, for example, ranks suppliers and factories by country risk, length of business relationship, and previous audit findings. While this may give the impression that the company is addressing its biggest problems, it hides that poor supplier practices may still persist – having simply been shifted around as suppliers focus their attention only to those areas under audit scrutiny while ignoring other areas in a constantly moving game of 'cat-and-mouse' obfuscation driven by the audit programme.

Accountability beyond the factory gates

Audits focus on the immediate factory environment, operations and employees while omitting large parts of the supply chain beyond the factory gates where the most egregious conditions often occur. Although many retailers and brand firms portray audits as 'comprehensive', they in fact only cover a small percentage of top-tier suppliers. Apple, for example, claimed in 2012 to be 'going deeper into the supply chain than any other company we know of', and 'reporting at a level of detail unparalleled in our industry'. And yet, Apple was only auditing a portion of its top Tier 1 suppliers and ignoring subcontract facilities. With one of the industry's longest established supply chain social and environmental responsibility (SER) audit programmes (implemented in 2001), HP openly admits in its 2013 CSR report, the gaps of its programme in evaluating 2nd and 3rd tier suppliers as well as 'non-production' call centres and labour agencies.

Because the path of social audits is typically built around a product supply chain rather than a 'labour supply chain', audits tend to exclude some of the most vulnerable workers within the supply chains.[57] While audits tend to look only at workers on the books, the most exploited workers tend to be employed through complex labour-subcontracting arrangements, and are several steps removed from producers' core workforces. As one of our informants described, 'there could always

---

[57] A labour supply chain 'consists of the sequence of employment relationships that a worker goes through in order to be deployed in a productive capacity'. See Allain, Crane, LeBaron, and Behbahani, *Forced Labour's Business Models*.

be another group of people'[58] who are not on the books. Our interviewees also explained that factory owners commonly rid their production facilities of agency or exploited workers during audits. In addition, because audits are 'snapshots' of labour practices on any given day, they do not encompass illegal practices that might have taken place prior to employment but nevertheless shape employment conditions. This is well illustrated by the major issue of debt bondage. As Apple describes in its 2012 annual report, labour recruiters within their supply chains often charge enormous recruitment fees, which workers then become indebted to pay back, and those debts are used to maximise exploitation and immobility during work. To its credit, Apple claims to have refunded $17 million in recruitment fees since 2008 alone,[59] but most companies have been far less responsive – their audit programmes overlook the issue.

These gaps in the audit mechanism are significant not just because of the large portions of the supply chain that are not measured (and thus, hardly 'managed'), but also because they illustrate how the pathway of most audits is structured in such a way as to conceal, rather than bring to light, some of the worst abuses in the supply chain.

Uneven measurement of labour and environmental problems

Audits conceal problems by the narrow range of company operations investigated but also by their uneven treatment of labour and environmental issue areas. Whereas 'people' (for example, employee wellbeing, worker conditions) are hard to quantify, environmental concerns translate more easily into business efficiency and financial gains. The case for competitive advantage in 'corporate greening' is well documented. Therefore, management systems and metrics to evaluate and benchmark environmental performance are more comprehensive than for labour[60] with the significant consequence of greater investment in environmental programmes at the growing cost of social welfare.

Environmental concerns are more closely considered a strategic investment issue with direct bearing on operational efficiency, access to capital (risk management) and the bottomline.[61] Using less materials and energy and cutting back on waste saves money and drives sales. General Electric, for example, set a 5-year $25 billion revenue target when it launched its Ecomagination clean technology environmental programme in 2005 and has since (according to the company's 2011 progress report) realised over '140 [environmental] products and solutions and $105 billion in revenue'. Social concerns on the other hand are addressed by a narrow set of traditional business metrics including total employee wage, benefits, and overtime expenditures that are too generalised to reveal the problems of actual working conditions and employee treatment essential to ensuring a healthy, safe, and stable workplace. Whereas companies increasingly consider environmental programmes an investment in greater efficiency, social programmes are treated as a business cost. Rather than weighing and measuring the anticipated rate of return of social programmes, companies generally accept that any returns are 'intangible' and adopt

---

[58] Personal communication with auditor, London, March 2013.
[59] See Apple, 'Supplier Responsibility: Empowering Workers', available at: {http://www.apple.com/supplier-responsibility/labor-and-human-rights/} accessed 1 September 2014.
[60] Ketty Kortelainen, 'Global supply chains and social requirements: Case studies of labour conditions auditing in the People's Republic of China', *Business Strategy and the Environment*, 17 (2008), p. 433.
[61] Michael Porter and Claas Van der Linde, 'Green and competitive: Ending the stalemate', *Harvard Business Review*, 75:5 (1995), pp. 120–51.

them on the basis that it is simply a 'necessary cost of doing (staying in) business'. The concealment of labour issues is further reinforced by how the audit pathway is built around the product supply chain and not the labour supply chain.[62] As a consequence, audits miss many of the growing labour problems such as factory owners increasingly 'handing off' their labour responsibilities by bringing in more and more temporary contracted workers for which they are not liable.

Company efforts reveal the uneven treatment of social and environmental concerns. Walmart, for instance, has pursued ambitious corporate environmental programs in recent years but not social compliance standards or social certifications. Consequently, they are winning awards for developing leading 'business solutions' with respect to climate change and renewable energy while at the same time representing a lightning rod for unfair labour practices (for example, forcing overtime, paying below minimum wage, and refusing employee requests to unionise). Nestlé also highlights the shortcomings of an audit-based compliance programme based on a bifurcated environmental and social approach. After proudly announcing the Fairtrade certification of its Kit Kat bar, Greenpeace launched a campaign against the product, calling it out for destroying Indonesian Orangutan habitat and rebranding it, 'ethical chocolate from environmentally destructive palm oil'. The highly successful campaign pointed the finger directly at Nestlé for trying to hide a damaging product behind its Fairtrade certification. In another, among a growing number of examples, the Lipton brand of tea have recently achieved the 'green' Rainforest Alliance Certification, in spite of the fact that there is widely reported child labour in its tea supply chain.[63] The bifurcation of environmental and social standards and audit programmes allows companies to work towards improvements in one sphere – typically the environmental sphere – while allowing problems to persist or even worsen in the other sphere.

Our site visit to Chun Wo Ho in 2012, also revealed firsthand how environmental concerns are treated differently than labour considerations due to the more tangible business value gains of environmental versus social programmes as embedded in certification and company audit programmes.[64] Chun Wo Ho is a medium-sized, family-owned and operated Chinese manufacturing company that makes accessories like trim and zippers for outdoor apparel and gear. Established in 1971, the company head office is in Hong Kong and its factories are located in and around nearby Shenzhen – China's fastest growing city and the home of Foxconn, one of Apple's biggest suppliers. Products produced include: lace and trimming, cordage, twine and string (drawcords), braid and webbing (including elasticised), and imitation leather and plastic belts. The company supplies its products to a growing customer base including major brands such as Nike, North Face, Converse, Mammot, and Disney who then sell to consumers worldwide through multinational retail stores such as Costco, Target, and Walmart. Chun Wo Ho's corporate tag line is four words: 'Reliable. Responsible. Integrity. Value.' The company has recognised the role of responsible, sustainable business practice for ensuring access to top tier North American and European markets and customers. While recognising the importance of

---

[62] Allain, Crane, LeBaron and Behbahani, *Forced Labour's Business Models*.

[63] See, for example, Gethin Chamberlain, 'How Poverty Wages for Tea Pickers Fuel Indian's Trade in Child Slavery', *The Guardian*, available at: {http://www.theguardian.com/world/2013/jul/20/poverty-tea-pickers-india-child-slavery} accessed 20 July 2013.

[64] Jane Lister and Genevieve LeBaron, *Shopping for Sustainability at the Canton Fair: The Political Economy of Transnational Retail Governance in China*, Liu Institute for Global Issues Working Paper Series 12-001 (Vancouver: University of British Columbia, 2012).

both environmental and social sustainability, the company has put greater emphasis on and investment in its environmental programmes. As the company director, Jason Mak explained, 'there are cost incentives for environmental reforms but no real cost incentive to increase health and safety or labor conditions more broadly'.[65]

In June 2009, Chun Wo Ho signed onto the Swiss-based Bluesign environmental programme (the apparel industry's sustainable production standard) to 'provide performance products with a reduced environmental footprint'.[66] To meet the Bluesign commitment the company spent $1 million on environmental improvements, specifically on new production equipment and process upgrades. Workplace safety improvements were also identified for future expenditure but at $100,000. These comprised one-tenth the amount of the environmental systems. Efficiency gains and investment payback were not expected from social investments.

Chun Wo Ho made the decision to invest $1 million in environmental improvements after calculating that, through cost-savings due to eco-efficiencies and increased sales from achieving Bluesign certification, they would be able to recover the investment and generate new profit within three years. But Bluesign would also yield an additional and longer-term source of cost savings: the replacement of human labour with machinery. In the short term, Mak explained, the cost of environmental improvements would be met by shedding and raising labour productivity.

Narrow and bifurcated audit metrics do not capture this interacting dynamic of environmental and social conditions – that is, how the supplier was looking to recover the cost of its brand buyer's and consumers' rising environmental expectations through changes to its workforce. This has global regulatory significance as this emerging business strategy is not unique to Chun Wo Ho. Because corporate environmental sustainability yields more measurable direct business value than social improvements, there is a growing tendency to privilege the former over the latter. Canton Fair vendors we interviewed consistently noted that environmental standards and certifications were more feasible and desirable than social ones. This was particularly the case given their retail customers' tight turnaround times, fluctuating order sizes, and requirements for low-priced production, which they informed us they were meeting largely through increased overtime and agency workers (which tend to be lower paid and face severe restrictions on their ability to assert rights). As companies have linked their growth strategies to clean water and state-of-the-art eco-friendly technology, workers rights, conditions, and job security have at best, lagged, and at worst, been undermined.

A major 2012 study of CSR in the apparel industry clarifies this point.[67] Assessing management systems in four categories – Policies, Traceability & Transparency, Monitoring & Training, and Workers Rights – the study evaluated 300 apparel brands on their efforts to address child and forced labour in their supply chains. Firms currently championed for their corporate environmental initiatives – such as Walmart and Adidas – ranked among the lowest in social performance and efforts to address forced labour. In spite of its improvements around renewable energy, waste, and carbon, Walmart's traceability, transparency, and workers' rights records ranked

---

[65] Personal communication at Chun Wo Ho factory, Shenzen, China, 27 April 2012.
[66] Bluesign, 'About', available at: {http://www.bluesign.com} accessed 4 August 2014.
[67] Haley Wrinkle, Elin Eriksson, and Adrienne Lee, *Apparel Industry Trends: From Farm to Factory*, available at: {http://www.free2work.org/trends/apparel/Apparel-Industry-Trends-2012.pdf} accessed 4 August 2014.

among the lowest in protecting against child and forced labour. In an illuminating paradox, Walmart now requires the seafood it sells to come from fisheries that are Marine Stewardship Council certified (or equivalent, or part of a credible fishery improvement project). Yet, the company is consistently accused of negligence around forced and bonded labour found within seafood harvest and packaging within its supply chain, such as in *The Guardian* investigation discussed earlier.[68] In a similar vein, Adidas is currently celebrated as an environmental sustainability leader, but its Monitoring & Training record is so insufficient that on 14 March 2013, Penn State University suspended its apparel contract with the company.[69]

In summary, compliance audits purport to provide objective evidence and comprehensive 'data' through standardised metrics on the social and environmental conditions at supplier factories. However, as this section reveals audit information is politicised, partial, and largely held confidentially. Retailers aggregate supplier improvement metrics resulting in praise in some areas while egregious violations go unattended in others. Audits give the impression of transparency and accountability when, in fact, in many cases, they obscure and deflect attention from the most damaging corporate practices. As a global regulatory mechanism they are ultimately serving to legitimise the *status quo* – reinforcing rather than transforming the underlying problem of unbalanced capital accumulation through buyer-supplier supply chain power inequities.

The ineffectiveness of the audit regime as a mechanism of global corporate regulation is underpinned, as this section revealed, by how audits are designed and practiced. The audit regime, however, cannot be 'fixed' through institutional redesign. While companies can address certain weaknesses in their audit programmes to increase supplier compliance such as by adjusting penalties[70] and improving auditor skills,[71] these efforts constitute only a 'tweaking' or 'tinkering' at the edges. They do not get at the much deeper systemic problems embedded in the power dynamics and as André Broome and Joel Quirk stress (in this Special Issue): the political agenda underlying the regime.

Conclusion

As part of a broader trend towards supply chain benchmarking, the audit regime ultimately disguises a normative, market-based governance agenda in seemingly objective systems of verification and associated metrics. Among insiders, the audit regime is widely accepted to be ineffective in addressing the environmental and social problems endemic to the transnational retail business model. Yet, in the face of deficient

---

[68] International Labour Rights Forum and Warehouse Workers United, 'The Walmart Effect: Child and Worker Rights Violations at Narong Seafood's Thailand Model Shrimp Processing Factory', available at: {http://www.laborrights.org/publications/walmart-effect-child-and-worker-rights-violations-narong-seafood} accessed 5 November 2013. See also Walmart, 'Making Seafood More Sustainable', available at: {http://www.walmartgreenroom.com/2013/03/making-seafood-more-sustainable-one-fishery-at-a-time/?utm_source=feedburner&utm_medium=feed&utm_campaign=Feed%3A+walmartgreenroom%2FIJlQ+%28The+Green+Room%29} accessed 3 August 2013.

[69] Simon Van Zuylen-Wood, 'Penn State Cuts Off Adidas Over Sweatshop Complaint', available at: {http://www.phillymag.com/news/2013/03/14/penn-state-2/} accessed 4 August 2014.

[70] Erica Plambeck and Terry Taylor, 'Motivating Compliance with Labour and Environmental Standards', Working Paper No. 3176, Stanford Graduate School of Business (March 2015), available at: {http://www.gsb.stanford.edu/faculty-research/working-papers/supplier-evasion-buyers-audit-implications-motivating-compliance} accessed 10 April 2015.

[71] Kortelainen, 'Global supply chains and social requirements', p. 441.

legally binding systems of supply chain governance, states are adopting the audit regime to create the illusion of transnational corporate monitoring where there is in fact very little. As NGOs have developed an interest in the reputational and financial benefits of supply chain benchmarking, they too have helped to codify and neutralise corporations' poor environmental and social records, and have deepened the credibility and legitimacy of auditing. Briefly put, while the drivers motivating industry, NGOs, and states to institutionalise the audit regime vary, their increasing participation reflects an alignment of these actors' interests and ideologies in market-based global governance. The growing adoption of audits as a tool of global corporate governance is bolstering corporate power, interests, and influence over consumers and policymakers, and ultimately, is deepening corporations' power to make their own rules and norms, and to evaluate and report on their own performance and adherence to those rules.

The consequences for global corporate governance are significant. As noted in Section I, states and financial institutions increasingly rely on audit-based information to estimate and assess corporate responsibility. Furthermore, states are increasingly adopting audit-based approaches to fixing serious supply chain problems. But given the serious shortcomings of the audit system, the prospects for addressing the key flaws of the global production system – including forced labour and slavery, or environmental destruction, pollution, and chemical spills and dumping – seem increasingly limited. Rather than improving practices, recent initiatives ultimately reinforce the privileging of private interests over protecting the public good in the global regulatory sphere. Indeed, that growing adoption of audits is accelerating at the same time as audit ineffectiveness is becoming more widely recognised – including by brand and retail businesses themselves – suggests that supply chain benchmarking is much more about reifying and legitimating standard business aims and practice than driving actual environmental or social improvements.

While we have endeavoured to show that the weakness of audits as a tool for driving environmental and social improvements is rooted in audit design, power relations, and implementation, it is also notable that certain weaknesses that have been previously identified in 'governing the world at a distance' through benchmarking do seem to limit the capacity of the audit regime as a form of global corporate governance.[72] Audits bring 'a semblance of certainty, relative standing, and measurable temporal change to an otherwise unmanageably complex world',[73] but that certainty serves to stabilise, legitimise and reinforce endemic problems within supply chains. Enclosing authority within expert knowledge, the symbolic power generated through audits has helped companies to stave off more stringent forms of regulation or restrictions on their activities. As with other benchmarks, therefore, the audit regime must be seen as 'an exercise in symbolic power that can alter the cognitive understandings that international actors use to make sense of the world and to purposively act upon it'.[74]

---

[72] See Broome and Quirk, 'Governing the world at a distance'; Liam Clegg, 'Benchmarking and blame games: Exploring the contestation of the Millennium Development Goals'; Harrison and Sekalala, 'Addressing the compliance gap?'; Alexandra Homolar, 'Human security benchmarks: Governing human wellbeing at a distance'; Caroline Kuzemko, 'Climate change benchmarking: Constructing a sensible future?'; Porter, 'Global benchmarking networks'; Leonard Seabrooke and Ducan Wigan, 'How activists use benchmarks: Reformist and revolutionary benchmarks for global economic justice'; Sending and Lie, 'The limits of global authority': all of the articles listed here in *Review of International Studies*, 41:5 (2015), pp. 813–1010.

[73] Broome and Quirk, 'Governing the world at a distance'.

[74] Ibid.

Addressing the compliance gap? UN initiatives to benchmark the human rights performance of states and corporations

JAMES HARRISON AND SHARIFAH SEKALALA

Abstract. This article examines under what conditions benchmarking and associated measurement initiatives produced by UN human rights actors could, and should, play a role in promoting compliance with international human rights norms. It is organised around a comparative analysis of UN benchmarking initiatives for states and corporations. With regard to states, the article argues that ideological misgivings and technical limitations have so far triumphed over aspirations that indicators and benchmarks might play a significant role in increasing compliance with international human rights norms. With regard to corporations, we find that measuring human rights performance has been framed by the recent UN Guiding Principles on Business and Human Rights using a much more expansive and less quantitative set of benchmarks. These latter benchmarks do not appear to be creating conditions under which the human rights performance of corporations is effectively interrogated, and as a result there is a danger of superficial legitimation. Comparative analysis of these two initiatives reveals some of the tensions inherent in utilising benchmarking in transnational efforts to achieve human rights compliance. It also allows us to contribute to broader debates about the quantification of performance and its potential and limitations as a tool of global governance.

James Harrison is Associate Professor in the School of Law and Co-Director of the Centre for Human Rights in Practice at the University of Warwick. He has published widely in relation to human rights norms and standards and their application to multinational corporations, international economic laws and institutions, and public spending decisions. He can be contacted at: {J.Harrison.3@warwick.ac.uk}.

Sharifah Sekalala is Research Fellow in the School of Law at the University of Warwick. Her research focuses on human rights, legal practices, and global health. She has published articles in journals such as *Global Health Governance* (in press), and *Hague Journal on the Rule of Law and Contemporary Politics*. Her book *Solving Global Health Problems: Lessons from AIDS, Malaria and Tuberculosis* is forthcoming with Cambridge University Press.

Introduction

International human rights norms have, for many years, been studied as an important element of wider scholarly debates about international norm dynamics and political change. In terms of Martha Finnemore and Kathryn Sikkink's norm life cycle, international human rights norms have clearly passed through the early stages of norm emergence and a form of 'norm cascade' has occurred.[1] This norm cascade, in the sense it is used here, refers to states' *commitment* to human rights, through for

---

[1] Martha Finnemore and Kathryn Sikkink, *International Norm Dynamics and Political Change, International Dynamics and Political Change* (Cambridge: MIT Press, 1998).

instance ratifying human rights treaties and engaging explicitly in human rights discourse. Every state in the international community has ratified at least one human rights treaty. Early stages of norm emergence and acceptance are therefore considered to be no longer 'particularly interesting' when studying international human rights norms.[2] It is in the transition from commitment to compliance where the real interest lies; how to translate a set of human rights commitments on paper, to 'sustained behaviour and … practices that conform to … international human rights norms'; what has been termed 'rule-consistent behaviour'.[3]

The transition from commitment to compliance is where scholarship on human rights norms becomes very contested. For a number of years, more sceptical scholars have undertaken empirical (and mostly quantitative) studies to support arguments that commitment by states to treaties on paper does not lead to better human rights protection in practice.[4] In some cases, it has been argued that commitment may even be a smokescreen, behind which governments can commit even worse abuses.[5] More optimistic scholarship argues that there is evidence that commitment can and does drive compliance.[6] This evidence is found mostly in studies in relation to violations of various civil and political rights where commitment combined with 'a sustained spiral of increased information, pressure and sanctions, from linked internal and external actors' can be seen to lead to positive change in multiple places.[7]

While there is much that is contested about this research, there is widespread consensus on two issues that are of key importance to this article. One important finding that has emerged is that commitment to international human rights norms does not by itself lead to compliance. Bridging the gap between commitment and compliance is a complex process. Due to the UN's weak institutional mechanisms for the enforcement of human rights protection, it largely has to rely on dialogue and persuasion. This has led key UN bodies to utilise what scholars have termed *orchestration*; the engagement of intermediaries to help with its monitoring and enforcement functions as a way of increasing compliance with human rights treaties.[8] There is widespread support for the idea that international and domestic civil society organisations are critical intermediaries in situations where states do improve their human rights performance.[9] This is because international and domestic NGOs are vital actors in sustaining the various techniques of dialogue and persuasion that might

---

[2] Thomas Risse, Stephen C. Ropp, and Kathryn Sikkink, *The Persistent Power of Human Rights* (Cambridge: Cambridge University Press, 2013). This can be compared with the approach taken in Risse, Ropp, and Sikkink, *The Power of Human Rights: International Norms and Domestic Change* (Cambridge: Cambridge University Press, 1999).

[3] Risse et al., *The Persistent Power of Human Rights*, p. 10.

[4] See, for example, Emilie Haffner-Burton, *Making Human Rights a Reality* (Princeton: Princeton University Press, 2013), ch. 5; and Risse et al., *The Persistent Power of Human Rights*, chs 2 and 3 for a review of other key articles.

[5] See, for example, Emilie Haffner-Burton, 'Sticks and stones: Naming and shaming the human rights enforcement problem', *International Organisation*, 62:4 (2008), pp. 689–716.

[6] See, for example, Beth A. Simmons, *Mobilizing for Human Rights: International Law in Domestic Politics* (Cambridge: Cambridge University Press, 2009); Christopher Fariss, 'Respect for human rights has improved over time: Modelling the changing standard of accountability', *American Political Science Review*, 108:2 (2014).

[7] Risse et al., *The Persistent Power of Human Rights*, p. 278.

[8] Xinyaun Dai, 'Orchestrating monitoring: the optimal adaptation of international organisations', in Abbott et al. (eds), *International Organisations as Orchestrators* (Cambridge, Cambridge University Press, 2015), p. 142.

[9] See, for example, Xinyaun Dai, 'The "compliance gap" and the efficacy of international human rights institutions', in Risse et al., *The Persistent Power of Human Rights*, p. 98; Haffner-Burton, *Making Human Rights a Reality*, ch. 5.

potentially lead to compliance with international human rights norms.[10] Their role is therefore scrutinised carefully in the case studies presented in this article.

A second important issue relates to the inherent problems and limitations of measurement of human rights performance in terms of both its *scope* and *methodological shortcomings*. In terms of its scope, most of the measurement that has taken place relates to compliance with a fairly narrow range of civil and political rights by state actors. There is very little systematic empirical research in relation to the majority of human rights norms, most notably economic, social, and cultural rights.[11] Nor is there significant research in relation to non-state actors, most importantly multinational corporations, where the literature has largely been concerned with the conceptual issue of who has human rights responsibilities, rather than how, and whether, those responsibilities are meaningfully enacted.[12] But recent scholarship has argued that 'the relevance of human rights is no longer regularly challenged within the business sector'.[13] As a result, the move from commitment to compliance in relation to corporations has started to become the subject of serious scholarly debate. This article contributes to broadening the scope of the measurement literature by examining a broader set of rights, (with a particular focus on economic, social and cultural rights) and by comparing the dynamics of the transition from commitment to compliance with regard to states and multinational corporations.

In terms of *methodothological shortcomings*, there is acknowledgement from most scholars working in the field that there are limitations in the research that attempts to measure the links between commitment and compliance, for a variety of reasons. Obtaining reliable data sources in an environment where perpetrators are likely to hide abuses and those collecting data have limited resources is problematic, and leads to questions over the reliability of data that is collected.[14] Most major sources of human rights data are compiled by a limited group of Western organisations and governments, and this may undermine the credibility of the measurement process in the eyes of some actors.[15] When data is aggregated, compared and ranked, problems of reliability and comparability of different data sources are exacerbated, and questions are raised about the degree to which we can draw meaningful and robust conclusions about the performance of individual countries and the human rights system as a whole.[16] We therefore recognise that methodological approaches in any human rights measurement activity, including those examined in this article, require careful interrogation.

This article engages with the above debates about both compliance and measurement by analysing an under-explored aspect of the human rights measurement phenomenon. It examines under what conditions benchmarking and associated measurement initiatives *within the UN system* could and should play a role in promoting compliance with international human rights norms. It is organised around a comparative analysis of two cases studies. First, we analyse measurement and benchmarking initiatives promoted by key UN actors in relation to *states'* human rights performance, with a particular focus on economic, social, and cultural rights. Second, we analyse a measurement and benchmarking initiative in relation to

---

[10] Michael Barnett and Martha Finnemore, *Rules for the World: International Organizations in Global Politics* (New York: Cornell University Press 2004), pp. 5–6.

[11] Haffner-Burton, *Making Human Rights a Reality*, p. 78.

[12] David Jason Karp, 'The location of international human rights practices: What is human rights practice?', *Review of International Studies*, 39:4 (2013), pp. 969–92.

[13] Risse et al., *The Persistent Power of Human Rights*, p. 238.

[14] Haffner-Burton, *Making Human Rights a Reality*, p. 84.

[15] Ibid.

[16] Risse et al., *The Persistent Power of Human Rights*, pp. 277–8

corporate human rights performance. This latter initiative flows directly from the UN Guiding Principles (UNGPs) created by John Ruggie, the UN Special Representative of the Secretary-General on Business and Human Rights. In both case studies, benchmarking emanates from measurement as a form of international governance (Type II, to use the typology of the editors of this Special Issue)[17] led by the relevant UN actors in relation to compliance with human rights norms. But in both cases, the subjects of the measurement process play a key role, as measurement is embedded within self-reporting processes whereby it is states and corporations respectively who have primary responsibility for providing the empirical data.

The method employed in both case studies is the same. First, we review official documentation and statements produced by key UN actors to understand the motivations and expectations of the measurement processes and benchmarking efforts that have been developed. We then present our own empirical analyses of those processes in order to evaluate the extent to which expectations have been realised. Throughout we engage with relevant scholarly literature that supports or contests our claims and those of relevant UN actors. We conclude the article by drawing comparisons between the two case studies.

In our first case study on *states*, we find benchmarking practices operating within a system of self-reporting and very much tied to the creation of an extensive range of quantitative and qualitative indicators. We find tensions here between the desire for indicators and benchmarking to strengthen compliance with international human rights norms, and technical and ideological misgivings about the role of quantitative measurement and comparative assessment in a human rights context. Ultimately we find that this tension has been resolved in favour of a very limited role for indicators and benchmarking in the overall monitoring of states' human rights performance. But we also find that international and domestic civil society organisations have been an integral part of the measurement efforts that have occurred.

In our second case study on *corporations*, we find that there is a great discrepancy between the expectations of benchmarking and what has happened in practice. Ruggie has argued that the UNGPs themselves create a set of authoritative benchmarks against which progress in relation to corporate human rights performance can be assessed. Benchmarking is framed here in terms of a much more expansive and less quantitative process. A central component of the Guiding Principles is, again, self-reporting of human rights performance by corporations. We argue that the benchmarks associated with this self-reporting process are, to a far greater extent than for states, in danger of leading to superficial legitimation of corporate human rights performance, without evidence that substantive action has taken place. We find that a failure to realise the importance of domestic and international civil society organisations is a crucial flaw in the way the UN Guiding Principles have been operationalised.

States and human rights: Indictors and benchmarks

The development of indicators and benchmarks by UN Actors

Efforts to develop measurement tools in relation to international human rights norms within the UN human rights institutions started with a focus on state responsibilities

---

[17] André Broome and Joel Quirk, 'Governing the world at a distance: the practice of global benchmarking', *Review of International Studies*, 41:5 (2015), pp. 819–41.

for economic, social, and cultural rights (ESC rights). ESC rights were seen as marginalised within the UN system compared to civil and political rights. They did not have the same 'infrastructure for their advancement' (including a process for individual petition or even, for many years, a monitoring body), and were often viewed as 'non-justiciable' and lacking in normative strength.[18] There was a feeling that, in order to advance ESC rights, the international community would need a more robust measurement infrastructure using better statistical data in order to capture more accurately a state's progression towards long-term goals on issues like health and education.[19] Indicators were seen as an important tool of measurement in this endeavour that would serve a number of useful functions, including the measurement of an individual state's progress over time (benchmarking performance of a single state) and the generation of comparative data about the performance of different countries (comparative benchmarking between states).[20]

Throughout the early years of these calls for indicators and benchmarking, it was very much up to the states themselves to create and apply them. The UN Treaty Monitoring Bodies (TMBs) were only involved in reviewing the practice that developed.[21] This was the result of a light-touch approach embedded in a system where the TMBs require states to submit periodic reports (every four or five years). Once reports are received, the TMBs, with input from NGOs, may request additional information from the state delegation during meetings in Geneva. The committees then issue 'concluding observations' in which they highlight areas of concern and make concrete recommendations for action.[22] It was primarily through this procedure that states were asked to make use of the indicators and benchmarks they had developed.

By the turn of the second millennium, however, the Committee on Economic Social and Cultural Rights (CESCR), which was charged with monitoring these rights, was arguing that a state's failure to monitor could in itself amount to a 'violation' of human rights.[23] Over time, there was also recognition of the need to develop a more systematic approach to the development of indicators. The World Conference on Human Rights in Vienna in 1993 and the subsequent Vienna Conference Declaration and Programme of Action addressed this. It concluded that, 'to strengthen the enjoyment of economic, social and cultural rights, additional approaches should be examined, such as a *system of indicators* to measure progress in the realization of the rights set forth in the International Covenant on Economic, Social and Cultural Rights' (para. 98).

Attempts to create a systematic approach to the use of statistical information and indicators in the UN system can be officially dated from a request by the

---

[18] Ann Janette Rosga and Margaret L. Satterthwaite, 'The trust in indicators: Measuring human rights', *Berkeley Journal of International Law*, 27:2 (2009), pp. 253–315.

[19] Scott Leckie, 'Another step towards indivisibility: Identifying the key features of violations of economic social and cultural rights', *Human Rights Quarterly*, 20:1 (1998), pp. 81–124.

[20] UN Special Rapporteur Danilo Tὄrk, 'Realisation of Economic Social, and Cultural Rights', Progress Report UN Doc. E/CN.4/Sub.2/1990/19 (1990).

[21] Ann Janette Rosga and Margaret L. Satterthwaite, 'Measuring human rights: UN indicators in critical perspective', in K. Davis, A. Fisher, B. Kingsbury, and S. Engle Merry (eds), *Governance by Indicators: Global Power through Quantification and Rankings* (Oxford: Oxford University Press, 2013), p. 299. See also Committee on Economic, Social and Cultural Rights, 'General Comment No.1; Reporting by States Parties', UN Doc. E/1989/22, annex III (1989), p. 87 requesting states to develop their own benchmarks at para. 6.

[22] For a more detailed exposition of the roles of the Treaty Monitoring Bodies, their strengths and weaknesses and proposals for reform, see Michael O'Flaherty, 'Reform of the UN human rights treaty body system: Locating the Dublin statement', *Human Rights Law Review*, 10:2 (2010), pp. 319–35.

[23] Committee on Economic, Social and Cultural Rights, 'General Comment No. 14', para. 52.

inter-committee meeting of treaty bodies in June 2006.[24] The Office of the United Nations High Commissioner for Human Rights (OHCHR) then formally started work on a more coherent conceptual and methodological framework through which human rights indicators could be created and utilised for measuring the performance of states in complying with their human rights obligations.[25] This work culminated in the production of a report in 2012 entitled 'Human Rights Indicators: A Guide to Measurement and Implementation'.[26] This report is the key text in setting out current thinking about the use of statistical information, indicators, and benchmarking within the UN human rights system.

Between forty and fifty 'structural, process and outcome indicators' are elaborated in the report in relation to a range of rights protected by core UN treaties. Structural indicators capture the commitment or intent of the state in meeting its human rights obligations (for example, relevant international human rights treaties signed and ratified by the state, National Plans of Action created, etc.). Process indicators relate to the efforts that states have made to transform their commitments into protection and promotion of relevant rights on the ground (for example, coverage of targeted population groups under public programmes; human rights complaints received and the proportion addressed). Outcome indicators capture the results in relation to relevant rights (for example, reported cases of miscarriage of justice and proportion of victims who received compensation within a reasonable time; educational attainments by targeted population group).[27]

Ultimately, the objective of this conceptual framework, the report argues, is to encourage 'a practical, transparent and structured approach for a comprehensive translation of the human rights standards into concrete, well-defined, contextually meaningful indicators that help in the promotion and implementation of human rights'.[28] However, the report also makes it clear that the choice of which indicators to apply in any given scenario will be up to the users 'in light of their objective and national context'.[29] This tension between comprehensiveness (between forty and fifty indicators for each human rights norm) and selectivity (states and other relevant actors can decide themselves which indicators to utilise) is symptomatic of the tensions at the heart of discussions about the proper place of measurement frameworks within human rights reporting processes. We return to this issue below.

Similar tensions are seen in OHCHR's perception of the way in which indicators are used within the monitoring of state compliance with international human rights norms. OHCHR claims that indicators have 'frequently' been used in state party reports to the UN TMBs, as well as in the concluding recommendations of these bodies to states. On the other hand, there are currently no suggestions by any UN human rights body that indicators will play anything other than a minor role in the practice of reporting by states and monitoring by UN TMBs. The OHCHR 2012 report states that 'qualitative and quasi-judicial assessments by independent human

---

[24] See United Nations Office of the High Commissioner for Human Rights (OHCHR), 'Report of the Chairpersons of the Human Rights Treaty Bodies on their Seventeenth Meeting', UN Doc. A/60/278 (2005); see also OHCHR, 'Report on Indicators for Monitoring Compliance with International Human Rights Instruments', UN Doc. HRI/MC/2006/7 (2006).

[25] OHCHR, 'Report on Indicators for Promoting and Monitoring the Implementation of Human Rights', UN Doc. HRI/MC/2008/3 (2008).

[26] OHCHR, 'Human Rights Indicators: A Guide to Measurement and Implementation', UN Doc. HR/PUB/12/5 (2012), p. 4.

[27] These indicators are all taken from OHCHR, 'Human Rights Indicators', p. 34.

[28] Ibid., p. 44.

[29] Ibid., p. 29

rights experts will continue to be the cornerstone of human rights assessment and monitoring, particularly for complex human rights issues'.[30]

There are a number of interlinked reasons for playing down the role that indicators can and should play in the state reporting process. The most prominent within the OHCHR report is that indicators are unable to reveal the full complexity of a state's human rights performance, even in relation to a specific right, and so remain a supportive and complementary aid to the process of making expert legal judgments about a state's human rights performance.[31] Such arguments are echoed by commentators who fear that the use of indicators will create an audit culture that is inconsistent with the fundamental ethos of human rights and their protection.[32] Other commentators have raised related concerns that over-reliance on indicators in making a judgment about the human rights performance of states will lead to penalising countries which are concerned about human rights and are trying to tackle the problems, as opposed to countries which are determined to suppress them (and so would be less likely to report on them).[33]

A second set of concerns arises from the resource constraints that many countries will face in measuring their own human rights performance. This is recognised within the OHCHR report, as well as in wider debates about the problems of the TMB process.[34] It is particularly problematic where evidence which must be collected to feed into the indicators requires extensive effort, for instance because it must be aggregated from a large statistically-representative sample of the population. This is very difficult for many developing countries to achieve, who, in addition to very limited budgets, will be affected by resource constraints such as poor rural infrastructure and lack of trained officials.[35]

An even more limited role is envisaged with respect to benchmarking in relation to these indicators. There is no intention of using human rights indicators to create a composite measurement framework of indicators to rank or benchmark countries according to their human rights performance. It is argued that the complexity of human rights is such that the methodological issues faced in such a ranking exercise would be monumental.[36] There are also more deep-seated concerns that it would be inappropriate because human rights are 'absolute standards that all societies have to strive towards; this aim cannot be diluted by creating relative performance benchmarks based on cross-country comparisons'.[37] Any comparison

---

[30] Ibid., p. 26.
[31] Ibid., p. 26.
[32] Rosga and Satterthwaite, 'Measuring Human Rights', p. 299.
[33] Philip Alston, 'Richard Lillich Memorial Lecture: Promoting the accountability of members of the new UN human rights council', *Journal of Transnational Law and Policy*, 15 (2005), pp. 49–96.
[34] OHCHR, 'Human Rights Indicators', pp. 85–6 and 103.
[35] UN Secretariat, 'Concept Paper on the High Commissioner's Proposal for a Unified Standing Treaty Body', UN Doc. HRI/MC/2006/2 (22 March 2006). See also Risse, Ropp, and Sikkink, *The Persistent Power of Human Rights*, for recognition of the need for capacity-building in order to achieve human rights compliance.
[36] OHCHR, 'Human Rights Indicators'. This issue also raises the question of whether this would be an acceptable use of resources for an already-stretched human rights community. See S. M. Engle, 'Measuring the world: Indicators, human rights, and global governance', *Current Anthropology*, 52:3 (2011), pp. 83–95. On the question of resources, see Carr Centre Project Report, 'Measuring Human Rights: Tracking Progress, Assessing Impact' (2005), available at: {http://www. hks.harvard.edu/cchrp/pdf/Measurement_2005Report.pdf} accessed 28 June 2014, p. 26.
[37] OHCHR, 'Human Rights Indicators', p. 30. Making a similar point, see R. J. Goldstein, 'The limitations of using quantitative data in studying human rights abuses', in Thomas B. Jabine and Richard P. Claude (eds), *Human Rights and Statistics: Getting the Record Straight* (Philadelphia: University of Pennsylvania Press, 1992), p. 40.

of state performance in the foreseeable future across countries would therefore be on the basis of comparison in relation to 'a few specific human rights standards at a time, such as the right to education or the right to life or aspects of these rights (e.g. literacy rates, reported disappearances), and not the entire gamut of human rights.'[38]

Despite the limited role envisaged for indicators and benchmarking within the UN TMB system, the use of indicators within the treaty monitoring process is envisaged by OHCHR as having a number of potential benefits. They could potentially streamline reporting processes, make them more transparent and effective, and enable NGOs to 'exercise more effective oversight' of human rights. This could lead to improved follow-up of the recommendations and concluding observations of the TMBs and other key human rights actors.[39] Could the conceptual framework of indicators created by OHCHR therefore play a significant role in increasing compliance with international human rights norms, or will the limitations, as envisaged by key UN actors, prevent this?

Self-reporting: Indicators and benchmarks in practice

The above analysis demonstrates the inherent tensions of OHCHR's approach to the use of indicators within the UN system: comprehensiveness versus selectivity, frequent use versus technical limitations that constrain usage, and aspirations that indicators might increase compliance with international human rights norms versus ideological misgivings about the appropriate place of comparative measurement within the international human rights framework. We undertook an empirical study of how human rights indicators and benchmarking are currently being utilised within the UN TMB system in order to understand how these tensions play out in reality.

Our study analysed the reporting process in relation to African countries (where resources are likely to be scarcest) reporting to the Committee for Economic Social and Cultural Rights (CESCR) (where calls for measurement tools have been strongest) since 2006 (when OHCHR started its work aimed at trialling the use of indicators). It was based on an analysis of all relevant materials available through the OHCHR database with respect to the reporting process of CESCR.[40] We selected four rights (the rights to health, education, work, and social security) in order to analyse how indicators in relation to these rights are being utilised. A framework of between forty and fifty structural, process, and outcome indicators have been created for all four of these rights by OHCHR. Indicators for the right to health and education are more likely to complement other measuring activity required of states, whereas the right to work and social security are more likely to require bespoke action. In particular this can be seen from the fact that health and education have seventeen and six Millennium-Development-Goal (MDG)-related indicators, respectively, whereas the right to work and social security have only two and one, respectively.[41] The full results of our analysis are available online.[42]

---

[38] OHCHR, 'Human Rights Indicators', p. 30.
[39] Ibid., p. 104.
[40] See OHCHR Treaty Bodies Database, available at: {http://tbinternet.ohchr.org/_layouts/treatybody external/TBSearch.aspx?Lang=en}.
[41] OHCHR 'Human Rights Indicators', p. 90f.
[42] See Centre for Human Rights in Practice, available at: {http://www2.warwick.ac.uk/fac/soc/law/research/ centres/chrp/projects/governance/benchmarking/}.

Twenty-one reports were produced by states over the period 2006–14.[43] In 86 per cent of cases, we found evidence that countries were reporting on all four of the rights that we were examining.[44] However, we found that states' use of indicators varied significantly between the different rights. Reporting on the right to social security was almost exclusively undertaken in narrative form, with only isolated examples of indicators being utilised.[45] Indicators were used a little more often in relation to the right to work, occurring in 24 per cent of cases, but were mostly limited to three (the number of employed persons in the private sector and public sector; the number of jobs created annually; and the overall unemployment rate). Indicators were used more extensively in relation to the rights to health and education, but were still generally limited to five or six core indicators.[46] Maternal health indicators were most frequently used, with 76 per cent of states utilising at least one of these during the reporting process.

Unsurprisingly, benchmarking of performance was never undertaken by states in relation to *other* states' performance. In our sample, 45 per cent of states benchmarked at least one indicator in relation to *their own* previous performance, but no state did this consistently. We were also interested in whether states were only benchmarking their performance when it was improving (a positive performance bias). However, this was not generally the case. For instance, in its report, Cameroon noted declines in the number of employers who observed health and safety requirements, granted leave to employees, encouraged the creation of unions or promoted the right to strike. Similarly, the Democratic Republic of Congo reported improvements in employment rates but declines in life expectancy, maternal birth mortality rates, and the number of students graduating.[47]

CESCR engaged with states over their use of indicators, or failure to use them, in a number of different ways. In its formal requests to states for more information once state reports had been submitted, they often asked countries to provide more data in relation to the OHCHR indicators. This could involve asking countries for new data, for more information on indicators already submitted, or to explain how legislative and other changes were affecting data that had been provided. However, there was no comprehensive attempt by CESCR to address all the indicators created in the OHCHR framework. Rather, they engaged sporadically and often did not request more information in relation to any indicators associated with the all the four rights analysed (see Table 1). In their Concluding Observations, CESCR are often in a situation where they are restricted to demanding better data from states to allow any form of measuring and benchmarking of individual

---

[43] This data is taken from the OHCHR website, available at: {http://tbinternet.ohchr.org/_layouts/treaty bodyexternal/TBSearch.aspx?Lang=en&TreatyID=9&DocTypeID=29}.

[44] Angola's state report (2008) was the only report submitted post-2006 which did not have a section on the right to social security, right to work and right to education. However in its most recent report in 2014 it reported on all four rights. There were also three state reports – Burundi (2013), Rwanda (2007) and Morocco (2013) – that were inaccessible.

[45] Egypt's 2010 report does include reference to the number of beneficiaries who received social security. CESCR, E.C.12.EGY.2-4 (22 December 2001).

[46] In relation to the right to education, the most frequently utilised indicators were illiteracy among people over 15 years; gender balance of the illiteracy rate; gross enrolment ratio for primary school education; gender balance of school enrolment and performance rates in general education. In relation to the right to health, the most frequently utilised indicators were the infant and child mortality rate; maternal mortality rate; percentage of deliveries attended by qualified staff; percentage of pregnant women who attend one antenatal appointment; and rate of prevalence of HIV and AIDS.

[47] Democratic Republic of Congo Report, E/C.12/COD/CO/4 (16 December 2009).

Country	Right to social security	Right to work	Right to education	Right to physical and mental health
Djibouti	Y	Y	Y	Y
Equatorial Guinea		Y	Y	Y
Egypt	Y	Y	Y	Y
Ethiopia		Y		Y
Gabon			Y	Y
Gambia	Y	Y		
Rwanda		Y		Y
Sudan	Y	Y		
Togo		Y		
Uganda				Y
Number of times further information requested/not requested	4/6	8/2	4/6	7/3

Table 1. *Illustrative table showing where CESCR has asked for more data in relation to indicators associated with the four rights analysed from 2012–14*

state performance to take place. For instance, in the concluding observations for Benin, the Committee questioned the lack of aggregated and updated statistical data, which prevented it from accurately assessing the state's implementation of the provisions of the Covenant.[48] Unsurprisingly, given the wariness expressed in the OHCHR report with respect to benchmarking *between* states, we found no examples of benchmarking by CESCR of one state's performance against another.

Finally, we investigated the role of NGOs in relation to the reporting process. Based on the materials catalogued in the OHCHR database on the monitoring process, NGOs provided written reports for 59 per cent of the cases in our study.[49] However, the number of NGOs who provided reports in relation to each state varied significantly. For instance, in Egypt, where there is a long tradition of civil society engagement, 11 groups submitted some form of shadow report, while in relation to Angola's 2008 report there was only a single factsheet from the Centre for Economic and Social Rights. In some states, such as Togo, a group of NGOs created a shadow coalition report.[50]

It is widely recognised that NGOs lack resources and training and this reduces their capacity to produce credible data.[51] However, there were numerous examples of NGOs utilising a range of formal OHCHR indicators and other statistical data to augment, challenge and disaggregate the information provided by states. For instance, Kenyan NGOs made extensive use of statistical data relating to OHCHR indicators on evidence of sexual violence against children, and to challenge the state's reporting on

---

[48] Benin. Concluding observations of the CESCR, CESCR, E.C12/BEN/CO//2 (9 June 2008) p. 2.

[49] CSS reports were received for Angola, Benin, Cameroon, DRC, Djibouti, Egypt, Ethiopia, Gabon, Gambia, Kenya, Mauritania, Rwanda, and Togo. Available from the OHCHR database at: {http://www.ohchr.org/_layouts/treatybodyexternal/TBSearch.aspx?Lang=en&TreatyID=9&DocTypeID=5}. The engagement of civil society is likely to have been much greater, as they will also have given oral evidence during the reporting process.

[50] Coalition PIDESC-Togo, 'Rapport de la coalition DESC-Togo sur différentes préoccupations des populations relatives aux droits reconnus par le PIDESC' (March 2013), available at: {http:www.ohchr.org/Treaties/CESCR/Shared%20Documents/TGO/INT_CESCR_NGO_TGO_13556_F.pdf}.

[51] OHCHR, 'Human Rights Indicators', p. 55.

this issue.[52] In the concluding report on Kenya, the TMB then challenged the government and suggested improvements on several issues raised by civil society, especially with regard to the high rate of sexual violence that had been reported by the NGOs.[53] But overall, engagement with indicators by NGOs was sporadic and there was no attempt to hold states accountable for reporting on the full range of indicators developed by the OHCHR. Instead, the examples we found demonstrate how indicators can be utilised strategically by NGOs to engage with, and to challenge, individual and specific arguments made by the state concerning their self-assessment of their human rights performance. These arguments can then be taken up by the TMB in their concluding observations. These concluding observations are then taken up by civil society organisations to campaign for better compliance with human rights norms by States in the future.[54]

Overall, our case study of African state reporting on economic, social, and cultural rights found that indicators and benchmarking play a limited role in the process. Indicators are utilised by most states in reporting to CESCR, but only in relation to a select number of indicators chosen by the states themselves. Instances of benchmarking are even more sporadic, and benchmarking is never used to compare one state's performance with that of another state. CESCR and NGOs also engage with indicators in a selective manner, focusing on a very limited number of indicators, and CESCR's role is often limited to demanding more information. Information provided in relation to indicators by NGOs is vital in that it is often an important mechanism allowing CESCR to challenge the state's account of its own human rights performance in its concluding observations.

Evaluating the role of benchmarking

Ultimately, ideological misgivings and concerns about technical limitations by key UN actors currently triumph over aspirations that indicators and benchmarks might increase compliance with international human rights norms. While efforts to use indicators in a systematic fashion are still in their infancy and there is some evidence of state engagement with indicators increasing slightly, embedded dynamics of the system make it difficult to envisage a significantly enhanced role in the future. The limited application of indicators and benchmarking stems, more than anything else, from the fact that states are largely left to decide how to engage with the framework created by OHCHR. Resource constraints mean that it would be impossible for many states to engage with the full gamut of indicators provided. But self-selection by states means that it will be extremely difficult to induce states to converge to any form of common practice and to systematically use these measurement tools as a mechanism

---

[52] Kenya, in its periodic report to CESCR, asserted that the stipulation of the 2001 Children's Act that no child shall be discriminated against on the basis of sex was evidence of its action in this area. But the report submitted by the Ipas Alliance utilised a variety of OHCHR indicators and other statistical data to challenge this assertion. This included data showing that up to 60 per cent of sexual offences were occurring against people younger than twenty years, and on the basis of a school-based study of 1,000 public secondary-school students in Nairobi, 40 per cent of sexually experienced secondary school girls reported that their first sexual contacts were unwanted. See Re: Supplementary information on the Kenya Government Report, scheduled for review by the Committee during its 41st session (3–21 November 2008), available at: {http://www.ohchr.org/Treaties/CESCR/Shared%20Documents/KEN/INT_CESCR_NGO_KEN_41_9388_E.pdf}.

[53] Kenya. Concluding report, E.C.12/KEN/CO/1 paras 14, 15, 22.

[54] For instance, the Centre for Economic and Social Rights for instance uses these concluding observations to create fact sheets for countries in order to create wider awareness about areas where socioeconomic rights are being abused.

for driving compliance with treaties. So the system largely avoids the tendency observed in other global governance initiatives; for benchmarking to have the effect of 'legitimating and promoting the recent histories and contemporary activities of Western states' and producing an overall 'performance score' that may have significant legitimacy issues.[55] At the same time, selective engagement by states with indicators, together with limited oversight and comparative analysis by the TMBs, combine to make it very unlikely that this limited form of benchmarking will significantly induce *compliance* of States in the way that has been observed with other benchmarking initiatives.

To the extent that there is any influence on state behaviour, the role of NGOs is important, in providing indicators that augment, challenge and/or disaggregate the information provided by the state.[56] Given the limited resources and in-country knowledge of the TMBs, this role is absolutely vital to the reporting process, as highlighted in the broader literature on the UN TMBs, and in the literature on civil society engagement with global governance mechanisms more generally.[57] NGOs are also then of critical importance in utilising the recommendations made by CESCR – including for state improvement in relation to particular indicators, and in relation to the collection of additional data – to push governments in particular countries to take action and improve the underlying human rights situation.[58] CESCR are in effect therefore utilising NGOs to play a key *orchestration* role.

Human rights and transnational corporations (TNCs)

The UN Guiding Principles on Business and Human Rights (UNGPs)

The UN Framework for addressing human rights and business has a very different history from the state-based system described above. The UN Framework and Guiding Principles on Business and Human Rights were created by United Nations Special Representative of the Secretary-General on Business and Human Rights, John Ruggie, over the last decade. Ruggie created the UN Framework (henceforth, 'the Framework') to deal with the failures and inadequacies of existing international governance mechanisms for effectively addressing human rights and business. Numerous multi-stakeholder, self-regulatory initiatives, and cross-border accountability mechanisms have attempted to hold businesses accountable for their human rights conduct. But these initiatives have, for the most part, exhibited sizeable weaknesses and failings and, overall, have left significant accountability gaps in terms of overall human rights protection.[59] The UN Framework and Guiding Principles sought to tackle this issue.

---

[55] André Broome and Joel Quirk, 'Governing the world at a distance'.

[56] Ibid., fn. 58.

[57] See, for example, Felice D. Gaer, 'Implementing international human rights norms: UN human rights treaty bodies and NGOs', *Journal of Human Rights*, 2:3 (2003), pp. 339–57. Analogous arguments are made about the role of NGOs in relation to other global governance initiatives. See, for example, Albert van Zyl, 'How civil society organizations close the accountability gap between transparency and accountability', *Governance*, 27:2 (2014), pp. 347–56.

[58] Philip Lynch and Ben Schokman, 'Taking human rights from the grassroots to Geneva … and back: Strengthening the relationship between UN treaty bodies and the NGOs', in M. C. Bassiouni and W. Schabas (eds), *New Challenges for the UN Treaty Body System and the Human Rights Council Procedures*? (Cambridge: Intersentia, 2011), p. 175.

[59] See James Harrison, 'Human rights and trans-national corporations: Establishing meaningful obligations', in Julio Faundez and Celine Tan (eds), *International Economic Law, Globalisation and Developing Countries* (Cheltenham: Edward Elgar, 2010).

The Framework addresses what states and business enterprises need to do to ensure business respect for human rights. It is augmented by the UN Guiding Principles on Business and Human Rights (the UNGPs), which work towards the implementation of the Framework and provide more detailed guidance as to how it should operate.[60] The Framework and UNGPs have been universally endorsed by the UN Human Rights Council, leading Ruggie to claim that they are 'the authoritative global reference point for business and human rights'.[61]

Instead of treaties, which states sign and ratify, Ruggie has instead created a series of principles, which require no formalised process of legitimisation by companies (other than a self-declaration of adoption in whatever form the corporation sees fit). The principles are intended to allow for measurement and comparison of the human rights performance of corporations, as well as the patchwork of different mechanisms for holding TNCs accountable for their human rights performance. In commentating on the UNGPs, Ruggie has repeatedly described them as 'benchmarks'. For instance:

> In short, where discord among states, businesses and civil society had paralyzed prior international efforts to encompass corporate conduct within the international human rights regime, there is now a common platform on which to build and authoritative benchmarks against which progress can be assessed.[62]

The UN Working Group on Business and Human Rights (the UN body which is responsible for overseeing the implementation of the UNGPs) has also commented that they comprise, *inter alia*, 'a set of *benchmarks* to enable stakeholders to assess and promote the respect of business, and its accountability, for human rights and engage with the issue'.[63]

Ruggie and the UN Working Group appear to be using the term 'benchmarking' here in a manner that is very different from the way the term is used within the state-based system, and is more analogous to its use in scholarship relating to business management. In his discourse on the genealogy of benchmarking, Anna Yeatman described benchmarking as a process whereby companies undertake a comparative exercise to ensure consistency of products between national companies with a view to improving the quality of products and processes.[64] Benchmarking was taken to mean the improvement of an entity's performance that is sought through comparison with an external point of reference or simply best practice.[65] Benchmarking, in this sense, is not necessarily linked to the development of statistical data or indicators but, rather,

---

[60] Human Rights Council. 'Council Holds Dialogue with Experts on Summary Executions, Independence of Judges and Lawyers, Transnational Corporations' (30 May 2011), available from: {http://www.ohchr.org/en/NewsEvents/Pages/DisplayNews.aspx?NewsID=11082&LangID=E} accessed 22 May 2015.

[61] United Nations News Centre, 'UN Human Rights Council Endorses Principles to Ensure Businesses Respect Human Rights' (2011), available at: {http://www.un.org/apps/news/story.asp?NewsID=38742#.VV3IavNwYdU} accessed 22 May 2015.

[62] John G. Ruggie, 'Progress in Corporate Accountability', Institute for Business and Human Rights (2013), available at: {http://www.ihrb.org/} accessed 25 April 2015. See also John G. Ruggie, *Just Business: Multinational Corporations and Human Rights* (New York, London: W. W. Norton & Company, 2013), p. 127, in which he again describes the UN Guiding Principles as a 'set of authoritative benchmarks'.

[63] United Nations Working Group on Business and Human Rights, 'Report of the Working Group on the Issue of Human Rights and Transnational Corporations and Other Business Enterprises', UN Doc. A/67/285 (2012), para. 6.

[64] Anna Yeatman, 'Interpreting contemporary contractualism', in Jonathan Boston (ed.), *The State Under Contract* (Wellington: Bridget Williams Books, 1978).

[65] On this, see, for example, Steven Casper and Bob Hancké, 'Global quality norms within national production regimes: ISO 9000 standards in the French and German car industries', *Organization Studies*, 20:6 (1999), pp. 961–85.

can be used in relation to a set of narrative principles that describe normative standards of conduct required by corporations.[66] Despite this fundamental difference in conception, Ruggie and the UN Working Group's aspirations for their 'benchmarks' are analogous to those of UN actors within the state-based system: that they will provide a mechanism by which we can assess or measure progress being made by corporations in complying with international human rights norms.

The Framework and UNGPs are based on three underlying principles: the state duty to protect against human rights abuses by third parties, including business; the corporate duty to respect human rights; and the need for more effective remedies. In this article, we focus particularly on the second principle of the Framework, the corporate duty to respect human rights and the UNGPs that elaborate on the content of that principle. This is where data collection and measurement of performance in relation to corporations takes place. At the heart of this principle is the process of human rights due diligence (HRDD) which is, in effect, a self-reporting process analogous to the state-based system of reporting discussed above. Ruggie defines HRDD as 'the steps a company must take to become aware of, prevent and address adverse human rights impacts'[67] and sets out his conception of what it entails in considerable detail in five of his Guiding Principles. These principles are effectively therefore the benchmarks that should be met in relation to the HRDD process.

The HRDD process requires that corporations consider their own potential and actual human rights impacts, and report on their findings. The principles include a prioritisation process so that companies undertake assessments where human rights risks are greatest (Principle 17); consultation with stakeholders and otherwise gathering evidence of potential or actual impacts (Principle 18); analysis of impacts in relation to recognised international human rights standards (Principle 18); integration of findings of HRDD within all relevant internal processes and functions of the corporation (Principle 19); monitoring or tracking the effectiveness of their response to impacts identified (Principle 20); and communicating with relevant stakeholders about how they are addressing their human rights impacts (Principle 21).[68]

What we have through these principles is the sketching out of a process of investigation and reporting whereby the company investigates, analyses, and reports on *its own* performance in relation to human rights norms and standards. The sense that this is essentially a self-reporting process is reinforced by the primary value that Ruggie sees in the process. He argues that HRDD can transform a company's relationship with human rights from one based primarily on 'naming and shaming' by external actors for corporate failures to respect human rights to one of 'knowing and showing' by companies regarding their own human rights performance.[69] The UNGPs therefore create a self-reporting process where the corporation is playing the same central role that the state is within the UN TMB system. But here we find a very different practice emerging, with a very different role for benchmarks and very different types of third party engagement with the system.

---

[66] See André Broome and Joel Quirk, 'Governing the world at a distance', where the same differentiation between conduct and result is made.

[67] Ruggie, 'Protect, respect and remedy', para. 56.

[68] On the similarity between this process and that of other human rights measurement processes see James Harrison, 'Human rights measurement: Reflections on current practice and future potential of human rights impact assessment', *Journal of Human Rights Practice*, 3:2 (2011), pp. 162–87.

[69] John G. Ruggie, 'Business and Human Rights: Further Steps Toward the Operationalization of the "Protect, Respect and Remedy" Framework', UN Doc. A/HRC/14/27 (2010), para. 80.

Self-reporting: UNGP benchmarks in practice

We reviewed the policies of 225 leading multinational corporations for this article. Companies were chosen on the basis of their size and their likelihood to be encountering human rights issues. We therefore included all members of the Equator Principles and Thun Group (banks), the International Petroleum Industry Environmental Conservation Association (oil and gas), and the International Council on Mining and Minerals (mining). Members of these associations were chosen because of the associations' corporate social responsibility (CSR) influence in sectors that have potentially large human rights impacts worldwide. We also included the twenty largest companies ranked in the Forbes 2000 list in the fields of apparel, technology, and food and beverage, mining, oil and gas and finance. These are all areas of business where human rights concerns are particularly prevalent.[70]

We searched the online resources of all the companies identified, using a variety of ways to ensure we found all human rights related commitments of the companies in question.[71] The full results of this review are available online.[72] We found that 46 of the firms had already explicitly adopted the Framework as part of their human rights policy, only three-and-a-half years after the publication of the UNGPs. This number can be expected to increase significantly in the coming years as a result of the adoption of the UNGPs in a range of international guidelines, standards, and frameworks. They are referenced or adopted, *inter alia,* in leading standards for responsible business conduct (for example, the OECD Guidelines for Multinational companies); key ethical investment indices (for example, the Dow Jones Sustainability Index and Microsoft's ethical investment company, Maplecroft); important social standards for investment (for example, the European Investment Bank's social and environmental performance standards); leading reporting initiatives (for example, the GRI Sustainability Reporting Guidelines); knowledge-sharing initiatives (for example, the Global Business Initiative of the Global Compact); and in sectoral codes and principles aimed at improving standards in particular industries (for example, the Voluntary Principles on Security and Human Rights, the International Code of Conduct of Private Security Sector Contractors, the Responsible Jewellery Council Codes of Practices, etc.).[73]

Ruggie and a range of other actors have been keen to emphasise the success of the UNGPs as authoritative benchmarks.[74] Success in this sense is measured in terms of

---

[70] No claim is made about the representitaveness of this sample. What is critical is that the selection is a relatively large sample of leading multinational corporations for whom human rights are likely to be a potentially significant issue, and therefore they are likely to have taken some action on this issue. As will become evident below, the main claims of the article relate to the subset of companies who have actually adopted/endorsed the UNGPS.

[71] Google was utilised for all searches. The first search was for 'Company Name + human rights'. This was supplemented by searches that involved the search terms 'Guiding Principles', 'Human Rights Due Diligence', 'Corporate Social Responsibility', and 'Sustainability'.

[72] See Centre for Human Rights in Practice, available at: {http://www2.warwick.ac.uk/fac/soc/law/research/centres/chrp/projects/governance/benchmarking/}.

[73] For a wide range of examples of adoption, reference, and/or implementation of the UNGPs by a range of different actors see the Business and Human Rights Resource Centre, 'UN Guiding Principles; Implementation – Tools and Examples', available at: {http://business-humanrights.org/en/un-guiding-principles/implementation-tools-examples} accessed 25 August 2014.

[74] Daan Schoemaker, 'Raising the Bar on Human Rights: What the Ruggie Principles Mean for Responsible Investors' (2011), available at: {http://www.sustainalytics.com/sites/ default/files/ruggie_principles_and_human_rights.pdf} accessed 27 June 2014; Susan A. Aaronson and Ian Higham, '"Re-righting business": John Ruggie, 'The Struggle to Develop International Human Rights Standards for Transnational Firms', available at: {http://www.gwu.edu/~iiep/assets/docs/papers/

the adoption of the UNGPs, or reference to them, by corporations themselves, as well as by a wide variety of other bodies that are involved in setting standards and/or monitoring human rights performance by TNCs. But this extensive adoption of and/or reference to the Framework and Guiding Principles by a wide range of actors is not, in itself, any real measure of their success in terms of actual improvement of human rights performance; it is only proof of *commitment* not *compliance*.[75] In fact, our review found that, in most instances, publicly available references to the UNGPs by businesses and other key stakeholders was extremely superficial. The vast majority of corporate adopters limit themselves to a general statement of commitment with no detailed information about how they are being implemented. In terms of other stakeholders, public information relating to the incorporation of the guidelines is generally extremely sparse, and can even be limited to a link on the website to the Ruggie Framework.[76]

When we focus on the narrower issue of HRDD, in our survey of companies, only 28 out of the 46 companies who commit to the Ruggie Framework (61 per cent) explicitly and unambiguously state that they are undertaking HRDD and of those companies, only one company (Nestlé) has any kind of report on the results of their HRDD process.[77] Many companies merely give vague statements that they are undertaking HRDD.[78] A number of other companies do provide some (mostly very limited) detail about their HRDD process, and some limited insights into the methods that they have adopted, but no actual reports of their human rights performance.[79] Where some detail is provided, there appear to be wide divergences in practice, with HRDD processes sometimes integrated into risk assessment processes, compliance monitoring, company audits, and/or impact assessment processes. The vast majority of practice is therefore occurring in private, with only the company itself, and, in some cases, an external evaluator chosen by the company, able to engage with the divergent practice that is emerging.[80]

Nestlé's report, produced in November 2013, represents the first time that a leading multinational company has publicly commented upon its HRDD process in its operations and supply chain. The report is mostly a commentary on the methodology of their HRDD process, with a summary of key findings and identification of key challenges. The report relates the results of the HRDD process that was undertaken with

Aaronson_IIEPWP2013-5.pdf} accessed 27 June 2014; Radu Mares (ed.), *The UN Guiding Principles on Business and Human Rights: Foundations and Implementation* (Leiden, Boston: Martinus Nijhoff Publishers, 2012), pp. 1–50.

[75] The terms commitment and compliance are utilised here in the sense utilised by Risse et al., *The Persistent Power of Human Rights*, p. 10. See the introduction to this article for a discussion of this.

[76] For instance, the Maplecroft website simply includes a (non-functional) link to the Ruggie Framework and the corresponding Guiding Principles without any indication of how they are utilised: see Maplecroft, 'Human Rights Due Diligence', available at: {https://maplecroft. com/services/human_rights_due_diligence/} accessed 22 May 2015.

[77] The database was last updated for the purposes of this article in December 2014. The Nestlé report is entitled, 'Talking the Human Rights Walk: Nestlé's Experience Assessing Human Rights Impacts in its Business Activities', available at: {http://www.nestle.com/asset-library/documents/ library/documents/corporate_social_responsibility/nestle-hria-white-paper.pdf} accessed 22 May 2015.

[78] See, for example, EcoPetrol, 'Management of Risks and Impacts', available at: {http://www.ecopetrol. com.co/english/contenido.aspx?catID=679&conID=77958} accessed 25 August 2014.

[79] See, for example, Barrack Gold Corporation, 'Human Rights Compliance Program', available at: {http://www.barrick.com/files/governance/Barrick-Human-Rights-Compliance-Program.pdf} accessed 22 May 2015.

[80] Identifying the same problems in relation to supply chain audits, see Genevieve LeBaron and Jane Lister, 'Benchmarking global supply chains: the power of the "ethical audit" regime', *Review of International Studies*, 41:5 (2015), pp. 905–24.

regard to its operations in seven countries where human rights risks were considered greatest. Even in these countries, the report makes clear that there were significant challenges, including the difficulties of adequately investigating all of Nestlé's operations in each country, engaging with particular stakeholder groups (women, the elderly, disabled people, migrants and children), and tackling the full range of issues demanded by a comprehensive human rights approach.

Evaluating the role of benchmarking

Our study of practice in relation to the UNGPs raises some significant concerns. HRDD is clearly an integral part of the UNGPs. Adoption of the UNGPs means that companies are required to undertake the self-reporting process that is represented by HRDD. However, only 61 per cent of companies who have publicly committed themselves to the UNGPs even identify that HRDD is taking place. The publicly available information about the actual results of HRDD is limited to one company report, summarising their experience of the process, rather than directly reporting its results. We know from the state-based system that for a country to report on its *overall* human rights performance, even to a single treaty body, is a complex process that is resource intensive and produces results of very variable quality.[81] Nestlé identified significant challenges in fundamental aspects of the HRDD process in their seven-country study. But multinational corporations are able to claim, as a result of adoption of the Ruggie Framework, that they are undertaking HRDD with respect to *all* their operations globally, without providing any public information about what they are doing (for instance, Microsoft claim: 'We assess the human rights impacts of all our operations on an ongoing basis').[82] As a result, there is significant danger of superficial legitimation of corporate human rights performance, without evidence that substantive action has taken place.[83]

On their own terms (that is, using benchmarking in the business management sense), the UNGPs appear to be failing as a set of benchmarks because they do not enshrine a process that, at the very least, creates conditions under which appropriate engagement of third-party actors can occur.[84] First, there is no equivalent to the TMBs in the corporate self-reporting process. The 'guardian body' of the UNGPs is the UN Working Group on Business and Human Rights, but they have not been given the powers or the resources to undertake any form of review of corporate performance. This means that the limited powers of the state-based system to *orchestrate* compliance with human rights norms have been entirely removed within the corporate one. Second, the UNGPs have not enshrined transparency as a fundamental 'benchmark' of the system.[85] As a result, and because the methodology and results of the assessments are not generally made publicly available, there is no

---

[81] Michael O'Flaherty, 'Reform of the UN human rights treaty body system: Locating the Dublin Statement', *Human Rights Law Review*, 10:2 (2010), pp. 319–35 (p. 321).

[82] Microsoft, 'Global Human Rights Statement', available at: {http://www.microsoft.com/en-gb/download/confirmation.aspx?id=41958} accessed 25 August 2014.

[83] Making the same point about superficial legitimation in relation to audits of corporate supply chains, see LeBaron and Lister, 'Benchmarking global supply chains'.

[84] On the importance of this, see, for example, Albert van Zyl, 'How civil society organizations close the gap between transparency and accountability', *Governance*, 27:2 (2014), pp. 347–356.

[85] Explaining this deficiency in the UNGPs, and setting out a number of important benefits of transparency, see James Harrison, 'Establishing a meaningful human rights due diligence process for corporations', in Faundez and Tan (eds), *International Law, Economic Globalisation and Developing Countries* (Cheltenham: Edward Elgar, 2010).

way in which individual assessments can be scrutinised by NGOs in the way that they are in the state-based system.[86] The vital role that we know NGOs play in inducing compliance where there are weak institutional mechanisms has therefore been lost. Nor is there even the possibility of peer-to-peer learning, which is professed to be a strength of other UN initiatives on business and human rights, since the majority of practice is occurring within individual corporate silos.[87]

At the same time, HRDD *is* increasingly being utilised by various actors to confer benefits on corporations who have undertaken it. Two examples are provided here. First, the Dow Jones Sustainability Index, which incorporates the Ruggie Framework and Guiding Principles, uses HRDD as part of its assessment process for determining sustainable investments, but only requires companies to make a brief statement pertaining to where they see HRDD as required in their operations.[88] Second, the US government requires HRDD to have taken place in relation to US companies that want to invest in Myanmar (Burma) with particular focus on the energy sector. In order to comply with the US Department of State Reporting Guidelines on responsible investment in Burma, corporations must disclose information on their policies and procedures, including those relating to human rights, worker rights, anti-corruption, land acquisition, grievance mechanisms, and the environment.[89] Only Coca Cola, of the companies who have so far reported, provided any detail of the HRDD process they undertook, and this was limited to a summary of findings that was insufficient to allow any kind of independent assessment of the robustness of the process undertaken.[90] These examples raise serious concerns. In the state-based system there was a refusal to engage in comparative benchmarking of performance between states, and yet in relation to corporations an analogous self-reporting process is used as the basis for the crudest possible form of ranking systems (companies that receive benefits, and those that do not from government bodies, ethical investment indices, etc.).

The failings of the HRDD process also highlight the problem of conceiving of benchmarking in the business management sense when addressing global governance challenges such as the human rights performance of corporations. Benchmarking efforts within self-reporting processes must create sufficient specificity to allow for confidence that corporate efforts are meaningful and worthy of comparison. Otherwise there is no way of gauging the extent to which corporations are actually moving towards compliance with human rights norms. Narrative principles that simply focus on process rather than provide guidance in relation to expected results do not seem well placed to capture these movements. Could more specific and

---

[86] On analogous problems of transparency within supply chain audits, see LeBaron and Lister, 'Benchmarking global supply chains'.

[87] Christian Voegtlin and Nicola M. Pless, 'Global governance: CSR and the role of the UN Global Compact', *Journal of Business Ethics*, 122:2 (2014), pp. 179–91, making the case for the Global Compact as a learning platform, and the importance of transparency in that context.

[88] See RobecoSAM, 'CSA Guide – RobecoSAM's Corporate Sustainability Assessment Methodology' (2014), available at: {http://www.sustainability-indices.com/images/corporate -sustainability-assessment-methodology-guidebook.pdf} accessed 28 June 2014, para. 3.2.

[89] See US Department of State, 'Responsible Investment Reporting Requirements – Final' (2013), available at: {http://www.humanrights.gov/wp-content/uploads/2013/05/Responsible-Investment-Reporting-Requirements-Final.pdf} accessed 22 May 2015, p. 3.

[90] See Embassy of the United States, Rangoon, Burma, '2013 Responsible Investment Reports', available at: {http://burma.usembassy.gov/reporting-requirements.html} accessed 28 June 2014, and The Coca-Cola Company, 'Myanmar Investment Due Diligence 2014 Annual Update Report' (2014) available at: {http://photos.state.gov/libraries/burma/895/pdf/TCCCDoSResponsibleInvestmentMMReport63014.pdf} accessed 28 June 2014, pp. 6–8.

quantitative indicators (such as those developed through the state-based reporting process) perform a role in directing corporations towards the type of human rights issues they should be measuring? Would benchmarking then stand on firmer foundations? Interestingly, it is clear from the limited information that is publicly available that companies who are making serious attempts to develop HRDD processes are creating their own sets of indicators against which to measure their performance.[91] However, the development of these indicators is happening in private, and the results are likely to be disparate and varied because insights about strengths and weaknesses will not be shared, and companies will make their own individual choices driven by internal factors. Such concerns are heightened by a survey of HRDD practice among multinational corporations, which found great variation between different companies.[92]

To avoid this fragmentation and disorder, might there be a case for relevant UN bodies re(inter)nationalising the production of indicators, as happened in the state-based monitoring system? Such a process should not be undertaken lightly. Would corporations acquiesce to a framework of indicators that constrained their own measurement activities? Furthermore, in terms of feasibility, the creation of a measurement framework by which to seriously and holistically interrogate corporate human rights performance would be a very complex and multifaceted process. The non-profit research and policy organisation NomoGaia has been a pioneer in undertaking human rights evaluation of corporate performance. They have produced four human rights impact assessments of corporate footprint projects, which have, at times, produced transformative results on the ground.[93] In undertaking these assessments they have produced a toolkit, which contains over 300 'topics' that need to be catalogued during the assessment process.[94] This type of measurement framework demonstrates the level of specificity and complexity involved in such endeavours, and should be seriously studied, as it could create the basis for a much more concrete shared understanding of what a self-reporting process should entail than the high level principles created by Ruggie that currently underpin corporate self-reporting practice.

The work by NomoGaia also reinforces the lessons that flow from the Nestlé study. It is clear that robust reporting on a corporation's human rights impacts has only developed to a stage where it is suited to an assessment of particular activities in specific physical locations.[95] Assessing the overall human rights performance of a company is a completely different scale of endeavour, for which tools need to be developed and tested iteratively over time. As it is, HRDD processes in relation to aggregated human rights performance by large multinational corporations that are not subjected to public scrutiny are feeding into important decisions by states, banks,

---

[91] See, for example, Barrack Gold Corporation, Human Rights Compliance Programme, available at: {http://www.barrick.com/files/governance/Barrick-Human-Rights-Compliance-Program.pdf} accessed 22 May 2015.

[92] See Human Rights Due Diligence, 'Summary of the results of the survey conducted by the German Global Compact Network and Econsense', p. 5, available at: {http://www.globalcompact.de/sites/default/files/themen/publikation/human_rights_due_diligence_summary_of_the_results_of_the_survey.pdf} accessed 22 May 2015.

[93] These are all available at: {www.nomogaia.org}. See also Salcito et al., 'Assessing corporate impacts in unstable contexts: a human rights perspective', *Environmental Impact Assessment Review*, 47 (2014), pp. 36–46 for a review of Paladin Energy's Kayelekera Uranium Project in Malawi.

[94] NomoGaia, 'Human Rights Impact Assessment: A Toolkit for Practitioners Conducting Corporate HRIAs', available at: {http://nomogaia.org/tools/} accessed 22 May 2015.4.

[95] Interview with the author of the Nomogaia methodology, and Lead Investigator on all four human rights impact assessment studies.

and a range of other actors about who is ethical or unethical, investing responsibly or irresponsibly. This is a dangerous phenomenon.

Conclusion

Our comparative examination of the aspirations and results of two different measurement initiatives within the UN system has demonstrated that benchmarking and related measurement activity are not currently being used by UN actors to seriously address 'the compliance gap' that has been identified in the academic scholarship. But the rationales for this limited impact are very different in our two case studies.

With respect to states, aspirations are limited as a result of an intentional decision to restrict the role of indicators and benchmarks because of ideological misgivings and technical concerns. The results are therefore largely to be expected. With respect to corporations, the aspirations for the benchmarks created by the UNGPs are much higher; they are intended to play a central role in increasing corporate compliance with human rights norms. The results are therefore of far greater concern. In the latter case, benchmarks, particularly in relation to HRDD, may be disguising rather than illuminating the steps that are required in order to translate commitment into compliance.

The absence of binding treaty commitments for corporations has meant that comparative analysis of the conditions under which states and corporations comply with international human rights norms has been very limited, albeit with some evidence this is changing.[96] Future academic research should involve much greater cross-fertilisation of ideas. For instance, the extensive academic research on the conditions under which states comply with international human rights norms highlights the importance of relevant UN bodies orchestrating domestic and international NGOs to help in the monitoring and enforcement of those norms. There is no reason to believe that this is any less true with respect to corporations. But our case study has demonstrated that NGOs are completely frozen out of the picture when it comes to the human rights self-reporting process undertaken by corporations in relation to the UNGPs. Comparative analysis clearly highlights this issue.

Future comparative scholarship should also consider the plurality of conceptions of 'benchmarking' that are in operation with respect to states and corporations. We have suggested that this plurality occurs as a result of the differing disciplinary origins of the term in the two case studies. A shared conceptual understanding of what the benchmarking process represents across the UN human rights system is an important bedrock for future endeavours in this field. Interestingly, the more quantitative conception of benchmarking, closely tied to the development of indicators, found within the state-based system has been the subject of much internal debate and scrutiny. This is driven, it seems, by fears of the 'quantitative' as failing to reveal the full complexity of human rights, as well as concern that such an approach is open to manipulation by actors who will try to 'play' the system. As a result, the role of 'the quantitative' is tightly circumscribed within a system that relies primarily on the judgment of experts expressed in narrative from.

On the other hand, the conception of benchmarks as narrative principles, unrelated to quantitative measurement, which is found in the corporate-based system has been subjected to very limited scrutiny within relevant scholarship. This is despite

---

[96] See, for example, Risse et al., *The Persistent Power of Human Rights*, which contains two chapters on business and human rights alongside four chapters on states and human rights.

the fact that, as this article has demonstrated, the latter conception of benchmarking has been extremely problematic. The different levels of scrutiny and attention accorded to these two conceptions surely in part relate to the nature of the human rights community itself – a community of social scientists more comfortable with narrative than numbers. Does this suggest the need to better engage with a broader range of academic expertise (mathematicians, statisticians, economists, etc.) who might be able to unlock the potential of benchmarking and other quantitative measurement processes and thereby allow them to contribute more seriously to reducing the 'compliance gap' with international human rights norms?

Suggesting more serious engagement with quantitative measurement raises more fundamental questions about the proper place of benchmarking efforts within broader global governance initiatives. Concerns have been raised in this Special Issue about how benchmarking initiatives can be utilised by transnational actors to make comparative evaluations of performance which increase public and private power over distant entities, without due attention being given to the complexity of the underlying phenomena being measured, and to the limitations inherent in the benchmarking process.[97] Such concerns appear particularly acute where the benchmarked entity has limited control over the benchmarks which are utilised, and how data is collected and/or interpreted; at the same time the consequences of poor performance for that entity are often potentially serious (for instance, as in relation to credit ratings produced by international credit rating agencies or the World Bank's Worldwide Governance Indicators).[98]

In both of our case studies, however, the power dynamics of the system operate very differently. Attempts to measure human rights performance occur within self-reporting initiatives, where very little power lies with the UN actors who have instigated the measurement efforts. Benchmarking, in this context, is a mechanism by which attempts are made to constrain states and corporations; to make them report in a manner that allows their performance to be more effectively scrutinised. The greatest concern in this scenario is that indicators and benchmarks are ineffectual (or even act as a smokescreen for poor performance) because they are not sufficiently demanding, nor are they consistently and transparently applied.

Attempts to develop and expand robust benchmarking practice (tied to indicators which are actually capable of measuring performance, and appropriate financial assistance for actors who need it) look less troubling in relation to these self-reporting initiatives. This is because they are tools that can potentially be used to pressurise relatively powerful actors (states and corporations) into taking important obligations more seriously, and make it harder for them to 'play' the system. Development and application of robust and transparent indicators and benchmarks also potentially encourages more meaningful third party engagement. All of this suggests that views on the appropriateness of particular benchmarking efforts should be informed as much by the dynamics of the system in which they are operating as by the inherent limitations of the benchmarks themselves to operate as legitimate and appropriate tools of global governance.

---

[97] See, for example, Alexandra Homolar, 'Human security benchmarks: governing human wellbeing at a distance', *Review of International Studies*, 41:5 (2015); Broome and Quirk, 'Governing the world at a distance'; and in a more nuanced sense, Ole Jacob Sending and Jon Harald Sande Lie, 'The limits of global authority: how the World Bank benchmarks economies in Ethiopia and Malawi', *Review of International Studies*, 41:5 (2015): all in this Special Issue.
[98] See Broome and Quirk, 'Governing the world at a distance'.

Benchmarking and blame games: Exploring the contestation of the Millennium Development Goals

LIAM CLEGG

Abstract. Benchmarking has long been a central component of the global development industry, with the most prominent recent initiative being the Millennium Development Goal (MDG) framework. However, within existing scholarship, the agent-level interactions surrounding the MDG framework remain under-explored. Here, on the back of an analysis of interactions that took place within and around key MDG review summits, I develop a typology to clarify the intersection of benchmarking and blame games. Overall, I demonstrate that despite the efforts of the MDG architects to insulate the initiative, blame games have permeated policymakers' engagements with the framework. Moreover, the content of these blame games have been carried over into the recently outlined Sustainable Development Goals (SDGs). A pattern of strategic clarification has seen the emergence within this follow-on SDG framework of more precise responsibilities on higher-income states to meet aid targets, and on lower-income states to meet governance reform targets. Given the deeply-embedded cleavages that were evident in UN review summits, similar blame games seem likely to follow the periodic evaluations within the SDGs' lifespan.

Liam Clegg is Lecturer in International Relations at the University of York. Liam's work on international organisations and global economic governance has been published in journals including *New Political Economy*, *Global Governance*, *Review of International Organization*, and *Review of International Political Economy*. Liam's first monograph (*Controlling the World Bank and IMF*) was published with Palgrave's IPE Series in 2013, and he is currently completing his second monograph (*Mortgaging Development: The World Bank and the Globalisation of Housing Finance*).

Introduction

The desire to boil the ungovernable complexity of the social world down to a series of manageable proxy-measures seems to be a ubiquitous feature of life in and around the modern state. In many cases the demand from state agencies for ever greater flows of data was initially linked to practices of war-making, with reliable information needed to guide the movement of standing armies and supplies, and the levying of taxes to fund these operations.[1] Over time, with the expanding reach of governing structures into other policy areas, the creation of data flows has proliferated. It is now standard practice for states to collate indicators on a wide range of demographic trends, forms of economic activity, the performance of education and healthcare systems, and many other areas.[2]

---

[1] Max Weber, *Economy and Society* (Berkley: University of California Press, 1978), pp. 956–1005; James C. Scott, *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed* (New Haven: Yale University Press, 1998), pp. 11–33.

[2] Michael Power, *The Audit Society: Rituals of Verification* (Oxford: Oxford University Press, 1997).

And while the instruments of the modern state may have enjoyed something of a head start, international organisations are now working hard to reproduce this quantification of the social world on a global scale. Although transformations across other areas of global governance have been dramatic,[3] it is perhaps in the realm of development that benchmarking practices are at their most extensive.[4]

Through this article, I focus on the Millennium Development Goal framework, which represents arguably the most prominent attempt to benchmark global development. The Millennium Development Goals were launched at the end of 2001, with an unprecedented number of heads of state signalling their support through a period of high-profile United Nations summitry. The performance of the international community as a whole through the early 2000s was, it was suggested, to be assessed by evaluating the extent to which the lot of its lowest income members was improved. The MDGs were introduced as a means of assessing the progress achieved in this regard. With its eight Goals and associated indicators and time-bound targets, the framework served to establish a multidimensional vision of poverty and poverty reduction for the twenty-first century. The MDG initiative has been the subject of much analysis through its fifteen-year lifespan, with a number of works emerging as the Goals' 2015 census point is reached.[5] Here, I contribute to this literature by exploring the interactions that took place in and around a series of UN reviews of the MDGs.

Through an analysis of archival records and primary documents, I demonstrate that blame games have permeated these MDG-related processes. To conceptually unpack these blame games, I develop a typology that classifies the attribution of responsibility within benchmarking exercises that differentiates between their design and execution phases. In the case of the MDGs, the framework's architects sought to produce an

---

[3] See André Broome and Joel Quirk, 'Governing the world at a distance: the practice of global bench-marking'; James Harrison and Sharifah Sekalala, 'Addressing the compliance gap? UN initiatives to benchmark the human rights performance of states and corporations'; Alexandra Homolar, 'Human security benchmarks: Governing human wellbeing at a distance'; Caroline Kuzemko, 'Climate change benchmarking: Constructing a sensible future?'; Genevieve LeBaron and Jane Lister, 'Benchmarking global supply chains: the power of the "ethical audit" regime'; Tony Porter, 'Global benchmarking networks: the cases of disaster risk reduction and supply chains'; Leonard Seabrooke and Ducan Wigan, 'How activists use benchmarks: Reformist and revolutionary benchmarks for global economic justice'; Ole Jacob Sending and Jon Harald Sande Lie, 'The limits of global authority: World Bank benchmarks in Ethiopia and Malawi': all of the articles listed here in *Review of International Studies*, 41:5 (2015), this pp. 813–1010. Also, Lorenzo Fioramonti, *How Numbers Rule the World* (London: Zed, 2014); Lorenzo Fioramonti, 'A post-GDP world: Why it's time to end the tyranny of Gross Domestic Product', *Foreign Policy* (2 June 2015), available at: {http://foreignpolicy.com/2015/06/02/a-post-gdp-world/} accessed 8 June 2015.

[4] For detailed insights into the historical origins of benchmarking and statistics in the field of international development, see Hans Arndt, 'Economic development: a semantic history', *Economic Development and Cultural Change*, 29:3 (1981), pp. 457–66. A significant foundation on which modern practices have been built is widely believed to have been provided by Eugene Staley, *The World Economy in Transition* (New York: Council on Foreign Relations, 1939).

[5] Oxfam, *Failing to Deliver: The IMF and the Millennium Development Goals* (Washington: Oxfam, 2003); Dennis Garrity, 'Agroforestry and the achievement of the Millennium Development Goals', *Agroforestry Systems*, 61:1 (2004), pp. 5–17; Andy Haines and Andrew Cassels, 'Can the Millennium Development Goals be attained?', *British Journal of Medicine*, 329:7462 (2004), pp. 394–7; Shalendra Sharma, 'The promise of monterrey: Meeting the Millennium Development Goals', *World Policy Journal*, 21:3 (2004), pp. 51–66; Suzanne Akiyama, 'Millennium Development Goals', in Naonobu Minato (ed.), *New Approaches to Development and Changing Sector Issues* (Tokyo: Foundation for Advanced Studies in International Development, 2005), pp. 34–50; Tamar Gutner, 'When "doing good" does not: the IMF and the Millennium Development Goals', in Deborah Avant, Martha Finnemore, and Susan Sell (eds), *Who Governs the Globe?* (Cambridge: Cambridge University Press, 2010), pp. 266–91; David Hulme and James Scott, 'The political economy of the MDGs: Retrospect and prospect for the world's biggest promise', *New Political Economy*, 15:2 (2010), pp. 293–306; John McArthur, 'Own the goals: What the Millennium Development Goals have accomplished', *Foreign Affairs*, 92:2 (2013), pp. 152–63.

initiative that used diffuse lines of responsibility to limit the prevalence of blame games. Developing country governments' low level of integration of the Goals into domestic policy frameworks served to maintain this opacity. However, UN review summits through 2008 and 2013 saw the emergence of prominent attempts to attribute blame for poor performance. A clear cleavage emerged between, on the one hand, representatives of developed states who laid primary responsibility for under-achievement on the MDGs at the door of developing countries' institutional weaknesses, and, on the other, representatives of emerging and developing states who singled out insufficient flows of aid.[6] Through a process of strategic clarification, enhanced attention has been placed on these issues within the Sustainable Development Goal framework that is in the process of superceding the MDGs.[7] Overall, through the article, I address the under-examination of agent-level dynamics within the MDG-related literature,[8] and I offer initial reflections on the relationship between benchmarking and blame games to the emerging literature on the politics of global benchmarking.

In developing this line of analysis, the article proceeds through the following structure. In the opening section, I outline the scope of the conceptual and empirical contribution that is made through the article. I introduce the four-fold typology to clarify the intersection between benchmarking and blame games, and provide an overview of the primary analysis that underpins this contribution. Through the second section I provide a history of the Millennium Development Goals, outlining the capturing of the drafting process by a narrow group of UN technocrats, and their layering of diffuse lines of responsibility into the design of the MDG framework. In the third section, I review the content analysis of MDG summitry and associated policy framework documents, which uncovers the ongoing practices of blame shifting through UN fora. This dynamic has seen developed country representatives laying responsibility for poor performance at the door of corruption and low institutional capacity in developing countries, and developing countries identifying developed countries' unmet aid commitments as the key factor inhibiting progress. Strategic clarification has come as the MDGs have evolved into the Sustainable Development Goals, whose relatively open and inclusive drafting process has led to the incorporation of more precisely delineated responsibilities into this follow-on system. Through the concluding section of the article, I provide a brief recapitulation of the article's core themes and findings.

The exploration of blame games in the world of the MDGs that I provide in the next section focuses on state representatives' engagements with the MDG framework that were manifest through the 2008 and 2013 UN review summits of the initiative, and through a series of post-MDG launch Poverty Reduction Strategy Papers (PRSPs).

---

[6] Throughout the article, the terms 'developed country' and 'developing country' are used to refer to members and non-members of the Organisation for Economic Cooperation and Development respectively.

[7] At the time of writing, the content of the Sustainable Development Goals had been drafted by the Open Working Group and reviewed and approved at the close of the 68th Session of the United National General Assembly. The final sign-off on the SDGs will take place in September 2015, with the UN Secretary General and his Special Advisor on the SDGs predicting overall stability in framework content. See Liz Ford, 'Sustainable Development Goals', *The Guardian Online* (19 January 2015), available at: {http://www.theguardian.com/global-development/2015/jan/19/sustainable-development-goals-united-nations} accessed 26 March 2015. Throughout the article, I use the term Sustainable Development Goals to refer to the version approved at the close of the 68th Session, which included 17 Goals and around 160 related indicators.

[8] Amongst the large body of MDG-related literature considered below, patterns of engagement and resistance displayed by particular agents in particular fora are explored in most detail by Sakiko Fukuda-Parr, 'Are the MDGs a priority for development strategies and Aid programs? Only few are!', *International Poverty Center Working Paper Series*, No. 48 (2008), pp. 1–28; Sakiko Fukuda-Parr and David Hulme, 'International norm dynamics and "the end of poverty": Understanding the Millennium Development Goals (MDGs)', *Global Governance: A Review of Multilateralism and International Organizations*, 17:1 (2011), pp. 17–36.

This picture is, of course, partial; the network of relationships surrounding the MDG framework extended out beyond this selection of actors and arenas. International and domestic non-governmental organisations, in particular, remained closely hooked in to the framework, and contributed to the emergence of extended blame games that unfolded across multiple levels of governance.[9] While these extended blame games sit beyond the central scope of the article, they represent an important subject for future research. Indeed, given the ambitions of the SDG architects to use the post-2015 framework to push for enhanced domestic accountability processes as a means of improving developing-country government performance, these extended blame games look set to become increasingly important over time.

Benchmarking and blame shifting in global development

In the realm of global development, benchmarks come in many different shapes and sizes. In many quarters the GDP measure continues to reign supreme, and has come to be seen as an objective yardstick with which to assess the performance of developed and developing countries alike.[10] Through the mid-2000s, several prominent critical appraisals sought to extend the range of data captured by this narrowly-targeted indicator, with proposals to augment the GDP figure with additional measures pertaining to both environmental and social factors.[11] Beyond these economistically-centred frameworks, more self-consciously multidimensional frameworks have been crafted by agencies of various stripes. These frameworks have been created to try to ensure that the way in which we measure and promote development includes an explicit focus on education, health, and other measures of living standards. The initiatives that can be listed in this regard include the United Nations Development Programme's Human Development Index, the World Economic Forum's Gender Gap Index, and, the focus of this article, the Millennium Development Goals.

In the paragraphs following I outline the extension offered by this article to both the emerging literature on the politics of global benchmarking, and the MDG-related literature. The former body of work offers many valuable insights, exploring both the constitutive and relational aspects of benchmarking interventions. By analysing the intersection between benchmarking and blame shifting, I in particular engage with works that focus on the relational dimension. Within the MDG-related literature, there is a tendency toward exploring the constitutive impact of the framework in a way that can obscure the role of agent-led contestation. By exploring the detail of agent interactions surrounding the MDGs, I rebalance the focus of these works. As is noted in the introduction to this Special Issue, benchmarks represent inescapably political interventions. Here, I draw on the roadmap provided by the editors of the

---

[9] See, for example, Duncan Green, 'The power of numbers: Why the MDGs were flawed (and post-2015 Goals look set to go the same way)', *Oxfam Blogs* (14 August 2014) available at: {http://oxfamblogs.org/fp2p/the-power-of-numbers-why-the-mdgs-were-flawed-and-post2015-goals-look-set-to-go-the-same-way/} accessed 8 June 2015.

[10] The taken-for-grantedness of this conceptualisation amongst groups inside the World Bank, for example, is touched upon by several analyses. See Anthony Bebbington, Scott Guggenhein, Elizabeth Olson, and Michael Woolcock, 'Exploring social capital debates at the World Bank', *Journal of Development Studies*, 40:5 (2004), pp. 33–64; Liam Clegg, 'Our dream is a world full of poverty indicators: The US, the World Bank, and the power of numbers', *New Political Economy*, 15:4 (2010), pp. 473–92.

[11] Nicholas Stern, 'The economics of climate change', *American Economic Review*, 98:2 (2008), pp. 1–37; Joseph Stiglitz, Armatya Sen, and Jean-Paul Fitoussi, *Report by the Committee on the Measurement of Economic Performance and Social Progress* (Paris: INSEE, 2009). More recently, see Fioramonti, *How Numbers Rule the World*; Fioramonti, 'A post-GDP world'.

Special Issue (André Broome and Joel Quirk) to outline the contribution from this work on the MDGs.

The first set of properties associated with benchmarking that are flagged up by Broome and Quirk serve to highlight the constitutive power of these systems. These properties relate to the reification of particular normative standards into observable and seemingly objective categorisations, and the commensuration of discrete units into a common evaluative metric.[12] By providing an authoritative ideal-type template that embodies a set of shorthand assumptions about how to understand and address a policy issue, the successful dissemination of a benchmark can serve to frame the way in which a community of relevant actors perceive significant aspects of their social and political environments. In existing literature, it is relatively common for benchmarking systems to be criticised for the blind spots they can create in their rises to dominance. By reproducing existing power relations, so critics argue, there is a tendency for benchmarking initiatives to crowd out the perspectives and interests of more marginalised agents.

Prominent contributions to feminist and critical accounting scholarship offer analyses that progress along these lines. Literature within this area has, for example, sought to highlight the systematic devaluation of social reproduction that has been locked in through the rise to dominance of the UN System of National Accounts (SNA) as the core point of reference shaping statistical agencies' attempts to track economic activity. By excluding labour undertaken within the household from its sphere of calculation, the SNA has been said to marginalise forms of work predominantly undertaken by females in both the study of political economy and the practice of economic management.[13] In a similar vein, the Millennium Development Goals themselves have also been attacked for their totalising impact. The MDG framework has been criticised, *inter alia*, for embedding a neoliberal conceptualisation of development, and for rolling out a top-down, technocratic approach that leaves little space for substantive input from domestic populations.[14] Findings from this article go towards shedding some light on the constitutive side of the MDG process. In particular, by demonstrating that the integration of MDG indicators into domestic policy frameworks has remained modest, and that contestation of the Goals within reviews remains widespread, I demonstrate that the constitutive impact of the MDGs remains relatively limited in the examined fora. Discussion over the appropriateness of the Goals, and use of alternative measures of development, has continued to feature prominently. It is, however, toward the more relational elements of benchmarking interventions that the core engagement from this article is directed.

The second set of properties belonging to benchmarking systems that are flagged up by Broome and Quirk in their introduction to this Special Issue refer to the

[12] Broome and Quirk, 'Governing the world at a distance'.

[13] Marilyn Waring, *Counting for Nothing: What Men Value and What Women Are Worth* (Toronto: University of Toronto Press, 2004); Michèle Pujol, 'Into the margin', in Edith Kuiper and Jolande Sap (eds), *Out of the Margin: Feminist Perspectives on Economics* (London: Routledge, 1995), pp. 17–34.

[14] Gabay classifies the body of scholarship adopting this stand on the MDGs as 'the reductionists'. Clive Gabay, 'The Millennium Development Goals and ambitious developmental engineering', *Third World Quarterly*, 33:7 (2012), pp. 1253–4. Particular works singled out by Gabay include Ashwani Saith, 'From universal values to Millennium Development Goals: Lost in transition', *Development and Change*, 37:6 (2006), pp. 1167–99; Peggy Antrobus, 'MDGs: Most Distracting Gimmicks', *Convergence*, 38:3 (2005), pp. 49–52; Samir Amin, 'The Millennium Development Goals: a critique from the South', *Monthly Review*, 57:10 (2006), pp. 1–9. See also Sakiko Fukuda-Parr, Alicia Ely Yamin, and Joshua Greenstein, 'The power of numbers: a critical review of Millennium Development Goal targets for human development and human rights, *Journal of Human Development and Capabilities*, 15:3 (2014), pp. 105–17; Gita Sen and Avanti Mukherjee, 'No empowerment without rights, no rights without politics: Gender-equality, MDGs and the post-2015 development agenda', *Journal of Human Development and Capabilities*, 15:3 (2014), pp. 188–202.

relational aspects of these structures. The relational impact of benchmarking comes with what the editors refer to as the closing of the recursive loop. Through benchmarking interventions, value rankings of an aspect of the social world are made, and the recursive loop is completed as benchmarked agents adjust their subsequent behaviour.[15] In its most simplified version, a hierarchical and unidirectional relationship can be seen to exist between the benchmarker, who designs the framework and operationalises the symbolic judgement, and the benchmarkee, who receives the judgement and responds accordingly. However, as is outlined elsewhere in the Special Issue and in the wider literature on the politics of global benchmarking, the relational aspects of benchmarking interventions can take the form of complex and iterative interactions.[16] Guided by the interactions observed through MDG review summits and associated processes, and drawing on insights from both the existing literature on blame games and on the politics of benchmarking, the article will present a framework for exploring the intersection of blame games and benchmarking interventions.[17]

To gain analytic traction over the intersection of blame games and benchmarking interventions, it is useful to begin by exploring the issue of responsibility. Two distinct moments can be identified that shape the nature of the lines of responsibility that become embedded within benchmarking interventions. The first of these comes at the benchmarking design phase. At this phase, the intervention's architects can opt to identify discrete agents with responsibility for the achievement of outcomes, agents who will be rewarded or sanctioned in line with benchmarked indicators. Conversely, these architects can design a benchmark with less direct lines of responsibility. In such cases, rather than being driven by the application of direct material or normative sanctions, any subsequent behavioural change will result from more diffuse shifts in beliefs about appropriate standards.[18] The second of these moments comes through the implementation phase. By virtue of their socially negotiated nature, lines of responsibility in the world of benchmarking retain a fluidity that outlasts their initially designated structure. As is comprehensively catalogued in the work of Christopher Hood and others,[19] many

---

[15] Broome and Quirk, 'Governing the world at a distance'.

[16] Please see Broome and Quirk, 'Governing the world at a distance'; Harrison and Sekalala, 'Addressing the compliance gap?'; Homolar, 'Human security benchmarks'; Kuzemko, 'Climate change benchmarking'; LeBaron and Lister, 'Benchmarking global supply chains'; Porter, 'Global benchmarking networks'; Seabrooke and Wigan, 'How activists use benchmarks'; Sending and Lie, 'The limits of global authority': all of the articles listed here in *Review of International Studies*, 41:5 (2015), pp. 813–1010. Sakiko Fukuda-Parr, 'Global goals as a policy tool: Intended and unintended consequences', *Journal of Human Development and Capabilities* 15:3 (2014), pp. 118–31.

[17] It is suggested by Rhodes that by acknowledging an iterative relationship to exist between conceptual clarification and empirical investigation, 'organising perspectives' can be established to guide our understanding of the social world. See Rod Rhodes, 'From marketisation to diplomacy: It's the mix that matters', *Australian Journal of Public Administration*, 56:2 (1997), pp. 40–53. In this study, an initial review of interactions highlighted the prevalence of blame shifting in MDG review summits, and prompted the subsequent drawing together of the work of Christopher Hood and others on this topic with other literatures on the politics of global benchmarking.

[18] Across the field of International Studies, much literature has sought to analytically and empirically cut between behavioural change associated with these 'logic of consequence' – and 'logic of approrpiateness' – based drivers. See, for example, James March and Johan Olsen, 'Institutional perspectives on political institutions', *Governance*, 9:3 (1996), pp. 247–64; Harald Müller, 'Arguing, bargaining, and all that: Communicative action, rationalist theory, and the logic of appropriateness in International Relations', *European Journal of International Relations*, 10:3 (2004), pp. 395-435. These issues are not directly explored in this article, although there is an implicit assumption that, at different times and in different contexts, changes in behaviour brought about by benchmarking interventions can be mediated through either form of causal pathway.

[19] Christopher Hood, *The Blame Game: Spin, Bureaucracy, and Self-Preservation in Government* (New Jersey: Princeton University Press, 2011); Christopher Hood, 'What happens when transparency meets blame-avoidance?', *Public Management Review*, 9:2 (2007), pp. 191–210. Earlier work on the politics of

		Direct	Diffuse
Responsibility	Direct	Direct responsibility	Strategic blurring
by	Diffuse	Strategic clarification	Diffuse responsibility
design			

Figure 1. *Responsibility in global benchmarking*

strategies are available to agents with which to either associate or dissociate themselves from a particular benchmarking exercise.[20] Through the implementation phase, lines of responsibility can be either clarified or blurred.

Given the existence of these two potential starting points and two potential processes of evolution, a four-part matrix can be created to capture the intersection between blame games and benchmarking processes. This typology is presented in Figure 1. I will briefly discuss dynamics associated with each mode, paying particular attention to benchmarking interventions in the field of development.

The inclusion of direct responsibilities within benchmarking interventions can be expected where asymmetric relationships exist between benchmarker and benchmarkee that enable the former to lock in their preferences against the contrasting preferences of the latter, or where there is a sufficient level of consensus amongst relevant parties. The asymmetric version of this mode coheres closely with practices associated with the new public management agenda; by collating and publishing a range of performance-related data, benchmarking systems are structured so as to encourage the assessed units to 'up their game'.[21] The increasing application of this mode of benchmarking in the realm of development is particularly associated with the rise to dominance of neoliberal practices through the 1980s.[22] In such cases, even if the benchmarkee disputes the legitimacy of the framework content or application, they will nonetheless receive material or normative sanctions in the case of under-performance. In order for this mode to be achieved through benchmarking interventions that bring together more equal partners, it is necessary for all parties to continue to 'buy in' to the framework through its execution. Where direct lines of

blame include Kathleen McGraw, 'Avoiding blame: an experimental investigation of political excuses and justifications', *British Journal of Political Science*, 20:1 (1990), pp. 119–31; Kathleen McGraw and Clark Hubbard, 'Some of the people some of the time: Individual differences in acceptance of political accounts', in Diana Mutz, Paul Sniderman, and Richard Brody (eds), *Political Persuasion and Attitude Change* (Michigan: University of Michigan Press, 1996), pp. 145–70.

[20] Although he rarely uses the terms 'benchmark' and 'benchmarking', significant aspects of Hood's work reflects on interactions that are in nature similar to those focused on by Broome and Quirk. Hood's favoured terminology for benchmark-like structures is 'synecdoche', which refers to the proxy measurement that is used to gauge performance as a whole. See Gwyn Bevan and Christopher Hood, 'What's measured is what matters: Targets and gaming in the English public health care system', *Public Administration*, 84:3 (2006), pp. 520–1.

[21] Patrick Dunleavy, Helen Margetts, Simon Bastow, and Jane Tinkler, 'New public management is dead – long live digital-era governance', *Journal of Public Administration Research and Theory*, 16:3 (2006), pp. 467–94; Chris Lorenz, 'If you're so smart, why are you under surveillance? Universities, neoliberalism, and new public management', *Critical Inquiry*, 38:3 (2012), pp. 599–629.

[22] For an overview of this transformation, see John Toye, 'Changing perspectives in development economics', in Ha-Joon Chang (ed.), *Rethinking Development Economics* (London: Anthem, 2003), pp. 21–40. In a similar vein, Jacqueline Best's recent work on transformations in development practice suggests that the increased preoccupation with managing risk and benchmarking performance has been driven by a growing preoccupation with policy failure through the 1990s. See Jacqueline Best, *Governing Failure: Provisional Expertise and the Transformation of Development Finance* (Cambridge: Cambridge University Press, 2014).

responsibility are layered in to the framework at the moment of design and maintained through the process of execution, the benchmarking intervention overall can be characterised as one of 'direct responsibility'.

From the same starting point of direct responsibility by design, an alternative mode can be reached through more dynamic patterns of interaction between the agents brought together by a benchmarking intervention. Where evaluations flag up poor performance, blame games can kick in to play as benchmarkees work to evade material sanctions and preserve reputation. While sanctions and reputational loss can be hard to avoid,[23] the invocation of mitigating factors can be used to minimise these costs by laying responsibility for sub-optimal outcomes with exogenous events.[24] In addition, third party actors can be used by the benchmarkee in blame avoidance techniques. In the realm of development, the International Monetary Fund in particular is often singled out as being responsible for perceived performance failures in national economies operating under its loan arrangements.[25] Where direct lines of responsibility are layered in at the moment of design and successfully challenged through the process of execution, the benchmarking intervention overall can be characterised as one of 'strategic blurring'.

Turning now to the modes in which the starting point for the benchmarking intervention was one of diffuse lines of responsibility, the relationships surrounding the Millennium Development Goals as explored in the article will provide useful exemplars. Diffuse lines of responsibility may be layered in to a benchmarking intervention where the architects are located on the margins of a given policy area and so are unable to control the flow of normative and material sanctions, or where there is a low level of consensus over the identity of the actors with the predominant capability to shape the benchmarked outcome. In the case of the MDGs, a combination of these two factors shaped the incorporation of diffuse lines of responsibility in to the benchmarking system. In addition to fearing that the creation of direct responsibilities on state actors might reduce their willingness to sign up to the scheme, the MDG architects were working in a context in which beliefs over the actors and factors influencing development outcomes remain heavily contested.[26] As such, the location of responsibility for the achievement of MDG-related indicators remained underspecified. Outcomes in the realms of education, health, sanitation, and the environment were diffused across the 'international community'. In the realm of overseas development aid, the one area where a relatively precise group of relevant actors can be seen to exist, responsibility was diffused through the creation of an imprecisely specified target.[27]

---

[23] Patterns of stability and change in actors' perceptions in this regard are explored in Thomas Preston, 'Weathering the politics of responsibility and blame', in Arjen Boin, Allan McConnell, and Paul Hart (eds), *Governing after Crisis: The Politics of Investigation, Accountability and Learning* (Cambridge: Cambridge University Press, 2008), pp. 33–61.

[24] McGraw, 'Avoiding blame'.

[25] For detailed exploration of scapegoating and the IMF, see James Vreeland, *The IMF and Economic Development* (Cambridge: Cambridge University Press, 2003).

[26] See, for example, Frederick Cooper and Randall Packard (eds), *International Development and the Social Sciences: Essays on the History and Politics of Knowledge* (Los Angeles: University of California Press, 2008).

[27] Interestingly, it seems that other attempts to benchmark the 'big picture' of development achievements have sought to embed more precisely delineated lines of responsibility than occurred with the MDGs. The 2005 Paris Declaration and its later Accra and Busan manifestations, for example, consciously aimed to adopt this more precise mode. In contrast to the MDGs, where the reluctance of the framework's closed team of architects to scare state representatives away from the benchmarking system led them to avoid imposing clear lines of responsibility, the Paris and later declarations were the product of more open drafting processes with more direct state representative involvement. See Bernard Wood, Dorte

From this starting point, an overall mode of 'diffuse responsibility' will be achieved if there is an overall pattern of stability in the relationship between actors brought together by the intervention. Through the lifetime of the MDGs, prominent dynamics served to create such institutional inertia. As is analysed in the article, by largely eschewing MDG targets and indicators within their national development strategies, developing country governments were able to avoid establishing a more direct line of responsibility for these benchmarked outcomes. However, additional interactions that occurred within and around the UN review summits served to generate a momentum that, in fact, was followed by an overall pattern of 'strategic clarification' as the MDGs were superseded by the Sustainable Development Goals.

Strategic clarification in benchmarking systems can be driven by both pull- and push-related factors. With the former, the emergent benchmarkee works to clarify their line of responsibility to the benchmarked indicator; with the latter, third party actors work to identify the benchmarkee to whom a clarified line of responsibility is to be attached. Actors' desire to secure reputational enhancement constitutes a significant pull-related factor. Where a benchmarked indicator begins to signal signs of success, actors may attempt to take ownership of positive performance.[28] Actors' desire to compel others into behavioural change constitutes a significant push-related factor. By establishing a clearer relationship between an actor and a benchmarked indicator that is signalling a sign of failure, normative and material sanctions can be deployed to catalyse change in the emergent benchmarkee.[29]

With the MDGs, it is through these push-related dynamics that strategic clarification has over time become layered in to the framework. In particular, throughout the MDG review summits that have been analysed, consistent calls were voiced for clearer lines of responsibility to be drawn to two additional components of the benchmarking system. On the one hand, a significant proportion of developing country representatives made calls to layer in a more precise overseas development assistance target into the framework, to be met by their developed country counterparts; on the other, many developed country representatives made calls to layer in more precise targets in relation to domestic institutional reform and anti-corruption measures, with responsibility tied to their developing country counterparts. While these elements remained excluded from the content of the MDG framework, they have become incorporated into the expanded set of Sustainable Development Goals that have recently superseded the 2001–15 initiative.

Kabell, Nansozi Muwanga, and Francisco Sagasti, *Evaluation of the Implementation of the Paris Declaration* (Paris: OECD, 2005). Thanks to the *Review of International Studies* anonymous reviewer for drawing my attention to these points of comparison.

[28] Christopher Hood, 'Public service managerialism: Onwards and upwards, or "trobriand cricket" again?', *The Political Quarterly*, 72:3 (2001), pp. 300–9. Similar dynamics are explored in Ben Clift and Jim Tomlinson, 'Whatever happened to the balance of payments "problem"? The contingent (re)construction of British economic performance assessment', *British Journal of Politics & International Relations*, 10:4 (2008), p. 607–29.

[29] Several studies have explored the way in which the Organisation for Economic Cooperation and Development's Programme for International Student Assessment has been invoked in this manner. See Ruth Dixon, Christiane Arndt, Manuel Mullers, Jarmo Vakkuri, Kristina Engblom-Pelkkala, and Christopher Hood, 'A lever for improvement or a magnet for blame? Press and political responses to international educational rankings in four EU countries', *Public Administration*, 91:2 (2013), pp. 484–505; Johanna Ringarp and Martin Rothland, 'Is the grass always greener? The effect of the PISA results on education debates in Sweden and Germany', *European Educational Research Journal*, 9:3 (2010), pp. 422–30. For more general reflections on this form of blame shifting, see Christopher Hood, *The Blame Game*, pp. 67–89.

By adopting an extended view that considers the MDG framework and its transition into the SDGs, and noting that the push towards the creation of more direct lines of responsibility emerged in the examined UN review summits, through this article I suggest that the MDG benchmarking intervention can be characterised as following the mode of strategic clarification.

The empirical material contained here comes from an analysis of the archival records connected to UN summitry on the Millennium Development Goals, and of a series of national development plans. The archival records drawn upon come from the repository of speeches delivered at two major Millennium Development Goal progress reviews, held at the United Nations headquarters in September 2008 and September 2013. In total, 86 contributions were analysed.[30] In addition to evaluating the level of support displayed for the benchmarking system overall and its component parts, I have also recorded whether clear calls for refocusing or extending the initiative were delivered. Where such a call was issued, the primary reform being advocated was recorded.[31] The national development plans analysed consisted of a randomly selected sample of one-third of the Poverty Reduction Strategy Papers published by low income countries by the close of 2013.[32] Given that UN documentation surrounding the launch of the MDG framework suggested that PRSPs be used by developing countries to outline their pathways towards meeting the framework's targets and indicators, this represents an appropriate location at which to further probe the presence or absence of strategic clarification through the lifetime of the benchmarking initiative.[33] The finding that the extent of engagement with MDG indicators within PRSPs remains limited serves to confirm and extend the conclusions offered by Fukuda-Parr's evaluation.[34] Overall, by demonstrating that sustained contestation within review summits has been followed by strategic clarification through the SDGs, new light is shed on the intersection between blame games and this most prominent of benchmarking initiatives.

The empirical analysis presented in the sections following provides an overview of the blame games involving state representatives in the world of the Millennium Development Goals. The terrain of blame games and the MDGs stretches out beyond these parameters, to other actors and to other levels of governance. In domestic political constellations across many developing countries, local and national actors have attempted to tie blame to governing authorities for perceived underperformance in relation to particular Goals. Indeed, it has been suggested by Bergh *et al.* that 'a large part of the MDG success story seems to be the role played

---

[30] 75 of these were delivered by state representatives, eight from representatives of non-governmental organisations, and three from representatives of inter-governmental organisations.

[31] Single counting was used in order to enhance the comparability of the data collected; had multiple counting been used to track each of the suggested areas of reform, noise from representatives with wide-ranging agendas would have effectively drowned out the more focused contributions. For discussion of the operationalisation of quantitative coding techniques, see Carl Auerbach and Louise Silverstein, *Qualitative Data: An Introduction to Coding and Analysis* (New York: New York University Press, 2003).

[32] The selected countries used in this study (and the year in which the PRSP was published) are: Bangladesh (2012), Burundi (2006), Cambodia (2005), Ethiopia (2005), Kenya (2004), Liberia (2008), Madagascar (2007), Malawi (2012), Rwanda (2007), Sierra Leone (2005), Tajikistan (2009), and Uganda (2010).

[33] Indeed, PRSPs were used by Fukuda-Parr when assessing the extent to which the MDGs had been disseminated across developing country governments. See Fukuda-Parr, 'Are the MDGs a priority?'. In order to gauge the extent to which (aspects of) the MDG framework have been incorporated into the overarching developmental vision these PRSPs, I have concentrated analysis on the Executive Summary or equivalent section. These sections typically run to between fifteen and twenty pages in length.

[34] Fukuda-Parr, 'Are the MDGs a priority?'. The extension relates to the chronological coverage and individual cases reviewed.

by domestic politics',[35] with progress having been catalysed when non-governmental organisations and others have helped generate such strategic clarification. While acknowledging the importance of these additional sites of contestation and recognising the need for further study of the intersection of blame games and benchmarking through these networks, the sections following review the dynamics as manifest in the key UN review summits and the surrounding PRSP policy documentation.

The MDGs and diffuse responsibility by design

Authorship processes matter in the world of global governance. Strategically important actors can go to extraordinary lengths to ensure that documentation providing an institutionally endorsed perspective coheres with their own preferences and ideas.[36] Here, I explore the impact of the writing process behind the formation of the Millennium Development Goals on the designation of responsibility for the performance indicators within the framework. The drafting of the MDGs was a highly atypical production process at the United Nations. UN reports are often the product of extended drafting processes, through which a wide range of interested parties are provided with an opportunity to review and suggest amendments to a given text. This writing by committee can lead to the production of balanced documents whose contents are widely approved, but which lack a clear 'take home message'.[37] In contrast, the tightly controlled process surrounding the MDG initiative saw direct input limited to a small circle. In crafting a framework capable of generating rapid endorsement from across the whole body of UN membership, the content laid down by the MDG authors was designed to exclude clear lines of responsibility for the achievement of particular development related outcomes. By design, the MDG framework was a benchmarking intervention with diffuse lines of responsibility.

It was on 18 September 2000 that the 189 member states of the United Nations adopted the Millennium Declaration. The Declaration was the core outcome of the Millennium Summit earlier that month, at which the largest gathering of world leaders in history had articulated their collective commitment to the values and principles of the United Nations. The Summit had been called by the General Assembly to 'provide an opportunity to strengthen the role of the UN in meeting the challenges of the twenty-first century'.[38] In keeping with the unprecedented scale and extremely high profile of the event, the text of the Declaration established an ambitious vision.[39] The agenda that was laid out was broad ranging, detailing hopes for progress in issue areas including reform of the UN, peace and security, and

[35] Gina Bergh, Marta Foresti, Alina Rocha Menocal, and Leni Wild, *Building Governance into a Post-2015 Framework: Exploring Transparency and Accountability as an Entry Point* (London: Overseas Development Institute, 2012).

[36] Robert Hunter Wade, 'US hegemony and the World Bank: the battle over people and ideas', *Review of International Political Economy*, 9:2 (2002), pp. 215–43.

[37] The information about drafting processes was provided by a former member of staff from the UN Habitat programme, in a May 2012 interview with the author. A similar account is provided by Fukuda-Parr and Hulme, 'International norm dynamics'.

[38] UN, *Millennium Declaration* (New York: UN, 2000).

[39] For a comprehensive review of contributions to the Millennium Summit, see Jerome Glenn, Elizabeth Florescu, and Theodore Gordon, *Analysis of United Nations Millennium Summit Speeches* (Georgia: Army Environment Policy Institute, 2001).

poverty reduction. It was around 18 months after the adoption of the Declaration that the Millennium Development Goals were unveiled as the vehicle through which these development commitments would be operationalised.

To understand the emergence of the Millennium Development Goal framework, it is necessary to switch venues from the United Nations to the OECD.[40] The OECD was established in 1960, and throughout the Cold War functioned as something as a 'think tank of the West'. Throughout the 1990s, following the close of the ideological standoff between the US and USSR, the OECD's leadership sought to enhance the policy relevance of its work programmes as a means of securing the institution's future role.[41] It is in this context that the OECD extension of its engagement with issues surrounding aid and development occurred. As part of its efforts to improve the effectiveness of bilateral assistance and the level of harmonisation amongst donor states, the OECD had in the late 1990s established a series of International Development Goals on which to focus their efforts. In 1999, the profile of the International Development Goal benchmarking system was significantly increased when the OECD joined with the European Commission, IMF, UN, and World Bank to conduct a progress review. In June 2000, as final preparations for the Millennium Summit were underway, the collaborative report on *A Better World For All* was published. The structure and content of the International Development Goals would, in due course, come to form the basis of the MDGs.

The transfer of these ideas from the OECD-led initiative into the MDGs was facilitated by the contracting in of actors involved with the former into the process surrounding the drafting of the latter. Once the initial green light had been provided by through the Millennium Summit, the task of transforming a somewhat amorphous Declaration into an actionable benchmarking system was taken up by a handful of individuals in UNDP and the United Nations Secretary General's Office (UNSGO). Under the leadership of Mark Malloch-Brown of UNDP and Michael Doyle of UNSGO, staff from across the OECD, UN, and World Bank were brought together to form a small MDG committee. All members of the committee had been involved in the *A Better World For All* publication, and the preliminary version of the Millennium Development Goals was framed around the same seven pillars that had provided structure to the earlier iteration. Through the process an eighth goal, relating to the formation of a global partnership to foster development, was layered in to the benchmarking system.

The Millennium Development Goals were presented to the UN General Assembly in December 2001, where the benchmarking system was 'noted' without objection from the floor.[42] The MDG system is made up of three interlinked levels. At the broadest level are the Goals. Each Goal has between one and six targets, which in turn are monitored by between one and four indicators. In total, the MDG framework includes eight Goals, 21 targets, and 60 indicators. The eight Goals at the heart of the framework outline a commitment to: eradicate extreme poverty and hunger; achieve universal primary education; promote gender equality; reduce child

---

[40] The following narrative account draws together insights put forward in Hulme and Scott, 'Political economy of the MDGs'; Fukuda-Parr and Hulme, 'International norm dynamics'; and Huck-Ju Kwon and Eunju Kim, 'Poverty reduction and good governance: Examining the rationale of the Millennium Development Goals', *Development and Change*, ifirst (2014), pp. 1–23.

[41] Von Bogdandy and Goldmann, 'International public authority', pp. 77–8.

[42] Such noting represents a lower-order form of acceptance than 'approval', which requires a positive vote. Fukuda-Parr and Hulme, 'International norm dynamics', p. 26.

mortality; improve maternal health; combat HIV/AIDS and other diseases; ensure environmental sustainability, and; establish a global partnership for development.

In line with their differing professional affiliations, the team involved in drafting the Millennium Development Goal framework had consciously aimed to create a benchmark around which a wide range of agencies could coordinate their activities. Drawing in major international organisations and leaders of developed and developing countries alike, the MDGs were designed as a tool to 'express the resolve of the world's political leaders to … make the right to development a reality for everyone'.[43] In order to minimise the possibility of any constituent group wishing to dissociate themselves from the framework or stall its progress through the UN General Assembly, the drafting team structured the framework in such a manner as to avoid directly apportioning responsibility for the achievement of individual elements on particular actors. Goals One to Seven were linked to precisely defined targets and indicators, but responsibility for the ultimate achievement of these outcomes was placed rather diffusely at the door of all members of the international system. Goal Eight specifically aimed to encourage developed countries to adjust their aid, trade, and other externally oriented policies so as to foster a more development enhancing environment; however, unlike the earlier elements of the framework, no quantitative metrics were tied to this aspect. The MDG authors were particularly wary of including the 0.7 per cent of GDP aid target, for fear of alienating in particular the US and Japanese leadership from the initiative.[44]

Driven by this imminent conflict-avoiding dynamic, responsibility for the achievements of the development outcomes being monitored by the Millennium Development Goal framework was diffused across a wide range of actors. With no direct mechanisms in place to hook material resources or normative suasion to the achievement of the Goals, the framework lacked the necessary foundations through which to compel good performance through the exercise of precisely targeted sanctions.[45] As is outlined in the next section, this lack of focus on the discrete responsibilities of identified agents would come to be criticised through subsequent reviews of the Millennium Development Goal framework. It is to the evaluation of these dynamics that I now turn. Initially, disengagement from the MDGs within national development strategies served to reproduce the diffuse lines of responsibility that had been embedded at the design phase. Through the 2008 and 2013 review summits the foundations of strategic clarification began to be laid; foundations that, in due course, would come to be incorporated into the successor Sustainable Development Goals framework.

The MDGs and the emergence of strategic clarification

The context into which the Millennium Development Goal initiative was born did not augur well for the success of the benchmarking intervention. The immediate aftermath of the launch of the MDG framework was a turbulent time across the UN system, with the schism created by the 2003 Iraq invasion and subsequent

---

[43] UN, *Implementation of the United Nations Millennium Declaration* (New York: UN, 2002), p. 8.
[44] Fukuda-Parr and Hulme, 'International norm dynamics', p. 29.
[45] The situation is succinctly summarised by John McArthur, who in a recent evaluation noted that '[t]he MDGs were not born with a plan, a budget, or a specific mapping out of responsibilities … No single individual or organization is responsible for achieving the MDGs.' McArthur, 'Own the goals', p. 154.

investigation into corruption within the 'oil for food' programme creating significant reputational damage.[46] The legitimacy of the MDG framework itself was also initially challenged by actors dissatisfied with the closed nature of the drafting process through which the initiative was created,[47] although over time the Goals did come to attract a high degree of support from across the international development policymaking community. As is detailed later in the article, initially the lines of responsibility surrounding the framework remained diffuse, with developing country governments by and large eschewing MDG targets and indicators from their Poverty Reduction Strategy Papers. However, through the MDG review summits of 2008 and 2013, the beginnings of a push toward strategic clarification began to emerge. Through the successor Sustainable Development Goals, this process of strategic clarification has been consolidated.

Within the 12 Poverty Reduction Strategy Papers that were reviewed, all contained a broad commitment to the framework. These national development plans are variously presented as aiming towards 'freeing our people from poverty and achieving the targeted MDGs',[48] achieving 'poverty reduction linked to the MDGs',[49] and creating the structural foundations whose establishment 'are directly linked to the achievement of the MDGs'.[50] However, the extensity with which the MDG framework was used to frame national development plans varied significantly. At one end of the scale, in their development plan the Cambodian government consciously and directly mirrored the MDG framework. Indeed, focusing as it did on achieving the Cambodian Millennium Development Goals, the ends of development included in this Paper borrowed their title and core content from the UN framework.[51] At the other end, in several development plans the MDG framework remaining a marginal feature, attracting just one or two brief mentions.

Further insights into the lines of responsibility between developing country governments and MDGs can be gained by exploring the detail of the engagement with individual targets and indicators in PRSPs. In keeping with the prominence of its discursive commitment, the Cambodian government leads the field in outlining a commitment to address 27 of the individual indicators from the MDG framework.[52] These indicators are drawn from across the spectrum of the MDGs, with aspects of Goals One to Seven all represented. Other instances of a relatively high level of correspondence between the content of the MDG benchmarking system and individual national plans come with the Tajik and Bangladeshi PRSPs, through which 14 and 13 indicators have been incorporated respectively. At the other end of the scale, while the Liberian PRSP did contain a discursive

[46] These issues are identified by Mark Malloch-Brown, who led the UNDP throughout this period, as having played a major role in holding back the progress of the MDG initiative. Mark Malloch-Brown, *The Unfinished Global Revolution: The Limits of Nations and the Pursuit of a New Politics* (London: Allen Lane, 2011), pp. 162–70.

[47] Roberto Bissio, 'Civil society and the MDGs', *Development Policy Journal*, 3:1 (2003), pp. 151–60.

[48] Government of the People's Republic of Bangladesh, *National Strategy for Accelerated Poverty Reduction* (Dhaka: National Planning Commission, 2012), p. i.

[49] Government of the Republic of Kenya, *Investment Programme for the Economic Recovery Strategy and Employment Creation* (Nairobi: Government of the Republic of Kenya, 2004), p. 1.

[50] Government of the Republic of Tajikistan, *Poverty Reduction Strategy* (Dushanbe: Government of the Republic of Tajikistan, 2009), p. 30.

[51] Royal Government of Cambodia, *Poverty Reduction Strategy Paper* (Phnom Penh: Royal Government of Cambodia, 2006).

[52] In total, 44 indicators are associated with Goals One to Seven of the MDGs, which are the aspects of the framework that relate to developing country outcomes. As such, even the best performing PRSPs display only a modest level of engagement in absolute terms.
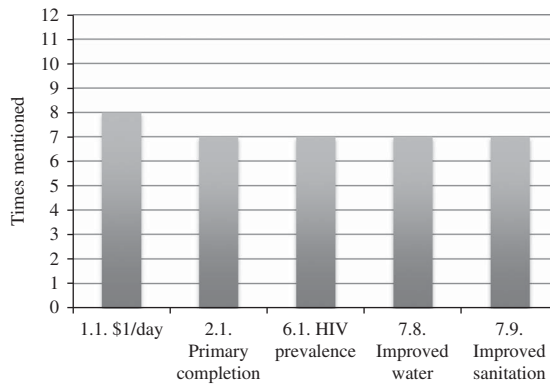
Source: Author's analysis of PRSP content.

Figure 2. *Most commonly appearing MDG indicators in PRSPs*

commitment to 'progressing toward the Millennium Development Goals',[53] no individual MDG indicators were included in the core focus of the Paper. The core focus of the Burundian, Kenyan, and Ugandan Papers included just one of the MDG indicators, coming in the form of the measures of the prevalence of HIV/AIDS, primary school enrolment, and the proportion of the population living on US $1 per day respectively.

Beneath this pattern of individual variation, common features of the PRSPs can be discerned that highlight where MDG indicators have – and have not – been integrated into national conceptualisations of the appropriate ends of development. The US $1 per day poverty benchmark was already a prominent feature of the development landscape by the launch of the MDG framework,[54] and as such it is perhaps unsurprising that this element was the most widely drawn upon within Poverty Reduction Strategy Papers (see Figure 2). Three-quarters of the Papers included this indicator as a core end to be achieved in national development programmes. Beyond this central marker for the first Millennium Development Goal, key measures from Goal Two regarding primary school completion, Goal Six regarding HIV prevalence, and Goal Seven regarding access to improved water and sanitation were present in the core aims of two-thirds of the Papers.

The establishment of clear lines of responsibility between developing country government and MDG indicator was, however, the exception rather than rule in the examined PRSPs. Beyond the five mentioned in Figure 1, none of the other sixty indicators from the MDG benchmarking system appeared in over half of the papers examined. In total, one quarter of the sixty failed to make it into any of the core aims of these PRSPs. Goal Five, which focuses on maternal health, remained the most overlooked.[55] Amongst the component indicators, measures relating to adolescent motherhood and unmet family planning requirements remained entirely absent, while indicators on births attended by a professional, the prevalence of contraceptive use, and antenatal visits appeared only in the Cambodian Paper. Goal Seven, which focuses on environmental sustainability, remained

[53] Government of the Republic of Liberia, *Poverty Reduction Strategy* (Monrovia: Government of the Republic of Liberia, 2008), p. 13.

[54] Kwon and Kim, 'Poverty reduction and good governance', pp. 2–3.

[55] Goal Five had a prevalence rate of under 10 per cent: from a potential 72 appearances, the six associated indicators were mentioned on just seven occasions.

the second most overlooked.[56] Indicators relating to the proportion of land mass covered by forest and the proportion of terrestrial and marine areas under official protection featured in just two Papers each. In addition, six indicators associated with Goal Seven remained entirely absent. Included amongst these were the measures of CO_2 emissions, consumption of ozone-depleting materials, and total water use.

Given that one-quarter of MDG indicators failed to feature at all in the reviewed PRSPs, and that the average prevalence rate was just 16 per cent,[57] it can be seen that by and large developing country governments remained disengaged from the benchmarking system through these documents. In the main, direct lines of responsibility between governments and Goals failed to emerge within PRSPs. Through the UN review summits of the MDGs, the dominant pattern of engagement from participants was to display a combination of effusive support for the framework in general terms, while raising points of contestation in relation to particular points of detail. It is through the latter that evidence of a push toward strategic clarification of the benchmarking system can begin to be seen, with developed and developing country representatives working to lay blame for under performance at each other's door.

At both the 2008 and 2013 events, virtually all participants presented strong praise for the MDG framework in their delivered statements. Examples include the following:

At the dawn of the new millennium … the Millennium Declaration was signed that provides a bold vision that … provides a crucial point of reference for measuring progress towards the establishment of a new world order that would be more equitable.[58]

The Millennium Development Goals were not a mere declaration. They are expressions of our common humanity, our common vision for a better world.[59]

At the beginning of the new millennium, our leaders came together … to express our shared commitment to reduce poverty and promote a more secure and prosperous world. The Millennium Development Goals express these shared commitments.[60]

However, underlying this superficial acceptance of the MDG framework as constituting a compelling vision for international development in the twenty-first century, significant discontent was displayed regarding the detail of the benchmarking system.

In total, around 85 per cent of the representations delivered at the UN review summits contained clear injunctions to extend the MDG framework.[61] Underlying these calls was a shared assumption that a higher degree of specificity was required in relation to the lines of responsibility surrounding the achievement of the benchmarked MDG outcomes. Overwhelmingly, developed country representatives shifted blame toward the high levels of corruption and poor institutional performance in developing

---

[56] Goal Seven had a prevalence rate of 15 per cent: from a potential 120 appearances, the ten associated indicators were mentioned on just 18 occasions.

[57] Goals One to Seven, which are the aspects of the MDG framework that relate to developing country outcomes, contain 44 associated targets and indicators. Across the 12 examined PRSPs, a 100 per cent prevalence rate would have required a total count of 528; the actual count of 86 equates to a prevalence rate of 16 per cent.

[58] Hamad Al-Thani, 'Statement from the State of Qatar at the High-Level Event of the United Nations on the Millennium Development Goals', UN General Assembly Hall, 25 September 2008, p. 2.

[59] Haim Divon, 'Statement from the State of Israel at the High-Level Event of the United Nations on the Millennium Development Goals', UN General Assembly Hall, 25 September 2008, p. 1.

[60] Julie Bishop, 'Statement from the State of Australia at the Special Event of the United Nations Towards Achieving the Millennium Development Goals', UN General Assembly Hall, 25 September 2013, p. 1.

[61] Of the 87 statements reviewed, 13 did not include a clear call for the MDG framework to be extended or refocused.

countries, while developing country governments shifted blame toward the unwillingness of developed country counterparts to live up to aid commitments.

The most commonly raised area for enhanced attention within the MDG framework related to the responsibility of developed country representatives to deliver appropriate volumes of aid flows. Discussions of Official Development Assistance (ODA) have an established lineage in UN forums; it was within the UN Conference on Trade and Development that, in the late 1960s, the much cited 0.7 per cent of GDP first emerged and gained traction.[62] The exclusion of this target from the MDG benchmarking system generated notable discontent. In total, this was the primary criticism offered by around one-third of participants at the 2008 and 2013 events. Through these contributions, inadequate levels of funding were highlighted as the core impediment to the more effective achievement of the ends of development as outlined in the MDG framework. The following calls are broadly representative:

All studies on the attainment of the MDGs have identified lack of finances as the main imperative … The imperative to reach the goal of 0.7 percent of Gross National Income on an urgent basis cannot be overstated.[63]

Inadequate financing … limits our capacity to implement MDG initiatives.[64]

[Poverty reduction] will remain elusive if the cooperating partners do not play their part and fulfil their role. It is sad to note that net official development assistance dropped by 4.7 percent in 2006 and a further 8.4 percent in 2007. Only five countries met or exceeded the 0.7 percent target.[65]

Amongst advocates of this extension, failures to reach the 0.7 per cent figure were commonly alluded to, and the fact that this target remained absent from the plethora of indicators within the benchmarking system attracted critical comment.[66] No OECD members were amongst the supporters of the extension of the MDG framework's focus on this issue.[67]

---

[62] For a detailed history of this target, see Michael Clemens and Todd Moss, 'Ghost of 0.7%: Origins and relevance of the international aid target', *International Journal of Development Issues*, 6:1 (2007), pp. 3–25. The evolution of the target within the UN can be traced back to the beginning of the first Decade of Development, and the General Assembly resolutions 1522 and 1711 of 1960 and 1961 respectively. These resolutions initially set the target at 1 per cent. Thanks to the anonymous reviewer for directing my attention to these roots.

[63] Palaniappan Chidambaram, 'Statement from the Government of India at the High-Level Event of the United Nations on the Millennium Development Goals', UN General Assembly Hall, 25 September 2008, p. 2.

[64] Frank Bainimarama, 'Statement from the Republic of Fiji at the Special Event of the United Nations Towards Achieving the Millennium Development Goals', UN General Assembly Hall, 25 September 2013, p. 1.

[65] Bradford Machila, 'Statement from the Zambian Delegation at the Special Event of the United Nations Towards Achieving the Millennium Development Goals', UN General Assembly Hall, 25 September 2008, p. 2.

[66] Arsenio Balisacan, 'Statement from the Government of the Philippines at the Special Event of the United Nations Towards Achieving the Millennium Development Goals', UN General Assembly Hall, 25 September 2013, p. 2; Wang Yi, 'Statement from the People's Republic of China at the High-Level Event of the United Nations on the Millennium Development Goals', UN General Assembly Hall, 25 September 2008, p. 5; Nancy Birdsall, 'Statement from Center for Global Development at the High-Level Event of the United Nations on the Millennium Development Goals', UN General Assembly Hall, 25 September 2008, p. 2; Celestino Migliore, 'Statement from the Holy See at the High-Level Event of the United Nations on the Millennium Development Goals', UN General Assembly Hall, 25 September 2008, p. 1.

[67] It is particularly interesting that China and India feature within this grouping, given that these states are increasingly major donors. More fine-grained analysis is required of the motivations behind these two particular interventions. While it remains plausible that this represents a case of blame avoidance through push-based strategic clarification, it is possible that other dynamics are at play. In aggregate, the

Where OECD country representatives at the UN review summits did advocate extension of the benchmarking system, their focus was firmly directed towards foregrounding the responsibilities of developing country governments in fostering effective progress. Attention on the role of domestic institutions in fostering development was relatively well established by the time of the launch of the MDGs, most notably at the World Bank.[68] At the UN events, strong advocacy for enhancing the focus on domestic institutional reform came from many representatives of developed country governments. Around half of the OECD representatives highlighted this issue as the key intervening variable in shaping successful outcomes. Comments from the US, Poland, and Sweden respectively capture the flavour of these injunctions:

Governance is the factor that best explains the performance of African countries ... The fight against corruption and entrenched interests remains a challenge in many places.[69]

It is of utmost importance that the community of recipients live up to the rules of good governance, and assumes responsibility for the initiatives designed to achieve the MDGs.[70]

We need democratic, effective, and accountable public institutions. We cannot fight poverty without fighting corruption ... Corruption undermines growth and prosperity, and an effective state and stable society. We ... have a responsibility to fight corruption, and we should make a special effort in the coming 800 days [leading to the MDG census point].[71]

From amongst the non-OECD members, the Rwandan representative was the sole representative to strongly endorse the call for a greater focus on domestic governance reform.[72]

As the MDGs drew towards their 2015 final census point, discussions over the form to be taken by their replacement became increasingly focused. It is through the creation of these Sustainable Development Goals that the process of strategic clarification, evident in the UN review summits outlined previously, was consolidated. In a marked contrast to the MDGs, the authorship process leading up to the launch of the SDGs has involved consultations with a wide range of state and non-state actors. As attempts to shape the SDGs reached their peak through a series of thematic workshops in late 2013 and early 2014, it was in fact suggested that 'a whole industry is now in gear to construct the global goals'.[73] Led by a thirty-member Open

---

interventions from the developing country grouping in UN review summits fall more straightforwardly under the rubric of push-based strategic clarification.

[68] For reviews of the embedding of this focus, see Weaver, *Hypocrisy Trap*; Liam Clegg, 'The governance of the World Bank', in Tony Payne and Nicola Phillips (eds), *Handbook of the International Political Economy of Governance* (Cheltenham: Edward Elgar, 2014), pp. 259–74.

[69] Henrietta Fore, 'Statement from the Government of the United States at the High-Level Event of the United Nations on the Millennium Development Goals', UN General Assembly Hall, 25 September 2008, p. 2.

[70] Ryszard Schnepf, 'Statement from the Republic of Poland at the High-Level Event of the United Nations on the Millennium Development Goals', UN General Assembly Hall, 25 September 2008, p. 4.

[71] Fredrick Reinfeldt, 'Statement from the Government of Sweden at the Special Event of the United Nations Towards Achieving the Millennium Development Goals', UN General Assembly Hall, 25 September 2013, p. 3.

[72] Indeed, there was notable resistance from other non-OECD members to such an extension. See Wen Jiabao, 'Statement from the People's Republic of China at the High-Level Event of the United Nations on the Millennium Development Goals', UN General Assembly Hall, 25 September 2008, p. 3; Yi, 'Special event', p. 4.

[73] Robert Wade, 'Current Thinking About Global Trade Policy', International Development at the LSE Blog (2013), available at: {http://blogs.lse.ac.uk/internationaldevelopment/2013/12/19/current-thinking-about-global-trade-policy/} accessed 21 April 2014.

Working Group,[74] 13 sessions were held between March 2013 and July 2014 to populate a 17-Goal template that had been generated through the earlier Rio + 20 UN Conference on Sustainable Development. The outcome from this process, which was released in July 2014, lists over 160 targets and indicators to be met by census points in 2020 and 2030.[75] Through the inclusion of more precisely specified obligations, the SDG framework is serving to enact a strategic clarification of the lines of responsibility surrounding the UN-centred global development benchmark (see Figure 1).

The disputes that were evident in the MDG review processes between developed and developing states over the lines of responsibility surrounding global poverty reduction have carried over into the Open Working Group discussions of SDG content. In the Open Working Group, discussions that sought to clarify and quantify the 'global partnership' needed to support development progress, while many of the contributions from developing countries pressed for the inclusion of the 0.7 per cent target, no support was forthcoming from developed country representatives.[76] In a similar vein, through the discussions of the role of good governance and the rule of law in fostering development developed country support was counterbalanced by developing country resistance.[77] Whereas in the case of the MDGs the tightly controlled drafting process led to the creation of a framework lacking clearly delineated lines of responsibility, the more open SDG drafting process has culminated in a markedly different endpoint. Covering the theme of global partnership, SDG 17 includes a direct call to developing countries to 'fully implement their commitments to provide 0.7 per cent GDP to developing countries'; covering the theme of accountable and inclusive institutions, SDG 16 calls for developing countries to 'substantially reduce corruption and bribery … and develop transparent institutions'.[78]

Viewed holistically, the interactions that have occurred within and around MDG review summits have, over time, created a push towards a strategic clarification of the benchmarking system. The diffuse lines of responsibility layered in to the benchmarking intervention by its architects were initially reinforced by developing country governments, through their low level of engagement with the Goals in national development strategies. However, with the launch of the Sustainable Development Goals, the pushes that emerged for more clearly delineated responsibilities on overseas development assistance and governance reform have recently been incorporated into the follow-on version of this most prominent of benchmarking initiatives.

---

[74] States were arranged into constituency groupings of between one and four members, with each grouping selecting a representative. For details of these groupings see UN Official Website, available at: {http://www.un.org/ga/search/view_doc.asp?symbol=A/67/L.48/Rev.1&Lang=E} accessed 22 August 2014.

[75] For the complete listing of these outcomes, see UN Sustainable Development Goals Official Website, available at: {http://sustainabledevelopment.un.org/content/documents/4518SDGs_FINAL_Proposal%20of%20OWG_19%20July%20at%201320hrsver3.pdf} accessed 22 August 2014. The document was reviewed at the close of the 68th Session of the United Nations General Assembly in September 2014. Final confirmation of the list of targets and indicators associated with the Sustainable Development Goals will be confirmed in September 2015.

[76] The only OECD country submission to the global partnership-focused section of the Open Working Group came failed to mention this target. See UN Sustainable Development Goals Official Website, available at: {http://sustainabledevelopment.un.org/index.php?menu=1680} accessed 18 February 2014.

[77] Compare, for example, the statement delivered on behalf of China, Indonesia, and Kazakhstan, with those from Canada, the Netherlands, and the UK, Canada, Israel, and the US, and France, Germany, and Switzerland. See the Sustainable Development Goals Official Website, available at: {https://sustainabledevelopment.un.org/topics/sdgs/group8} accessed 22 August 2014.

[78] See UN Sustainable Development Goals Official Website, available at: {http://sustainabledevelopment.un.org/content/documents/4518SDGs_FINAL_Proposal%20of%20OWG_19%20July%20at%201320hrsver3.pdf} accessed 22 August 2014.

Conclusion

Benchmarking is widely acknowledged to be an inherently political form of intervention. Having gained a high degree of prominence across the community of international development practitioners and scholars since their launch in 2001, the Millennium Development Goal benchmarking initiative has over the years attracted much comment and analysis. In the lead up to the final census year of 2015, this level of attention has increased further. Through this article, I have contributed to this MDG-related literature by presenting an empirically-focused exploration of agent-level dynamics surrounding the initiative. The insights generated through the paper are founded on new analysis of archival material from UN reviews of the MDGs undertaken in 2008 and 2013, and of the content of a series of national development plans as presented in Poverty Reduction Strategy Papers.

Through my analysis of this material, I outlined the existence of prominent blame games around the Millennium Development Goals. To explore this intersection between blame games and benchmarking in global development, I brought observations from this case together with existing literature on blame games and benchmarking. Through this process a four-part typology was created, which differentiated between benchmarking modes according to the nature of the lines of responsibility that were layered in to the system at the point of design, and whether there was stability or change in the nature of these lines of responsibility over time. Initially, the MDG architects incorporated diffuse lines of responsibility in to the framework. Challenges to this opacity began to emerge through UN reviews, with pressure emerging in particular for clearer lines of responsibility in the realms of overseas development assistance and domestic governance reform. With the superseding of the Millennium Development Goals with the Sustainable Development Goals, this push has been consolidated. As such, viewed holistically, the MDGs can be seen to have followed the mode of strategic clarification.

Beyond this embedding of calls for strategic clarifications initiated within MDG review processes, a number of additional important changes have occurred with the emergence of the Sustainable Development Goals. Perhaps most intriguingly, by extending the coverage of the framework from a limited subsection of lower income states to all members of the UN system, the SDGs will challenge an entrenched dichotomy between 'developed' countries to whom development prescriptions and yardsticks do not apply, and 'developing countries' to whom they do. When aligned with its inclusion of calls to prioritise the income growth of the poorest 40 per cent of domestic populations, SDG prescriptions can be seen to cut against the grain of contemporary distributional trends in many countries. As was the case with the MDGs, blame games and processes of contestation will determine which, if any, of the many targets and indicators associated with the SDGs gain traction across national and international policymaking processes.

The existing academic literature on the Millennium Development Goal framework provides valuable explorations of the power relations that were reflected in and (re)produced by this benchmarking initiative. While presenting important insights, the high level of abstraction within much of this literature served to underplay the role of agency and agent-level interactions in shaping the politics of the Millennium Development Goals.[79] Here, I have extended the MDG-related literature

---

[79] As outlined earlier in the article, this point of critique is most readily applicable to the 'reductionist' literature highlighted in Gabay, 'The Millennium Development Goals', pp. 1253–4.

by systematically exploring the agent-level blame games that have accompanied the MDGs. Through the examined interactions, developed-country representatives were found to have sheeted home responsibility for underperformance to their developing country counterparts, and developing country representatives were found very much to have returned the favour. With the incorporation of clarifications on aid- and governance reform-related benchmarks in the SDG framework, both sides' calls are being carried forward into the post-2015 initiative.

In important respects, the elite-level blame games that took place in and around the UN headquarters building represent a starting point in the politics of MDG and SDG global development benchmarking interventions. Throughout the lifetime of the MDGs, UN review summits provided a location in which strategic clarification of the lines of responsibility took place at dispersed intervals. Domestic political constellations across the global South provided locations in which parallel contests, undoubtedly, took place at a much greater level of frequency. These domestic blame games remained by and large off the radar of International Studies scholarship. With the SDGs' enhanced focus on governance reform, accountability, and transparency, it is likely that the domestic blame games surrounding the post-2015 benchmarking exercise will be played with equal or greater levels of intensity. In order to extend our understanding of the intersection between blame games and benchmarking in global development, it is important that these coming interactions are subjected to a more complete examination.

Climate change benchmarking: Constructing a sustainable future?

CAROLINE KUZEMKO

Abstract. This article analyses discourses on climate change and mitigation through the deconstruction of European Union (EU) rhetoric and practices on climate benchmarking. It critically examines the motivations behind climate benchmarking, the methods used to construct international benchmarks, and the reasons for variety in domestic compliance. Germany and the United Kingdom are analysed as cases where domestic politics drive very different reactions to the practice of climate mitigation, differences that have been largely hidden by the type of quantification that EU benchmarking involves. Through an exploration of the methods used to formulate climate benchmarks, the article demonstrates that these commitments have privileged certain responses over others, and thus helped to paint a picture of EU benchmarks as 'reformist' but not 'radical'. EU climate benchmarks often end up concealing more than they reveal, making it difficult to fully engage with the scale and complexity of the far-reaching domestic changes that are required in order to comply with agreed international benchmarks. The deficiencies of benchmarks as a mechanism for driving long-term sustainable change, and importantly discouraging harmful policies, may ultimately undermine their credibility as a means for governing climate change at a distance in the EU.

Caroline Kuzemko is part of an EPSRC funded project, 'Innovation and Governance (IGov)', based in the Energy Policy Group at the University of Exeter. She has research interests in energy and climate governance, (de)politicisation, and institutional change. She is a the (co-)convenor of two international, interdisciplinary academic networks: the PSA sponsored 'Anti-Politics and Depoliticisation Special Group' (APDSG) and the 'Political Economy of Energy in Europe and Russia' (PEEER). Caroline has recently published *The Energy Security-Climate Nexus: Institutional Change in the UK and Beyond* (2013 Palgrave Macmillan) and articles in the *Journal of European Public Policy*, and *Policy and Politics* (2014).

Introduction

For at least thirty years some international organisations (IOs), governments, and non-governmental organisations (NGOs) have worked hard to establish and diffuse new global norms in order to mitigate for anthropogenic climate change. A variety of speeches and texts associated with these organisations reveal arguments about the severity of implications for humankind if climate change is not mitigated, but arguably less understanding of the scale of the task that is being set nor of the degree to which fossil fuels are embedded within current economic models and political relations. Amongst international organisations some degree of agreement about how to proceed has been reached. There are a number of binding international climate targets including those agreed within the United Nations Framework on Climate Change Convention (UNFCCC) and the EU 20-20-20 agreement. There has been a proliferation of new climate agencies, committees, and international instruments,

including benchmarks, tasked with delivering these targets. The EU is taken here as an important actor, some say a leader, within such governance processes of trying to mitigate for climate change within severe temporal constraints. In this way their successes and failures are considered important in setting precedence of possibility but also because governance methods chosen and endorsed here may be followed elsewhere.

This article peers beneath these, at least rhetorically, committed climate discourses by exploring EU climate change benchmarking practices. Benchmarks are understood here as one important method utilised within the process of establishing new international climate norms and driving associated sustainable policy and behaviour changes. Climate benchmarks are the point of reference against which individual country's climate performance may be compared or assessed – climate targets agreed internationally, or in this case regionally, are used as the standard of measurement. There are those that have argued that climate mitigation targets should be too important for or above politics, often understood in terms of conflict, whilst others claim that politics gets in the way of successful climate mitigation and sustainable energy transitions.[1] In that benchmarks reflect decisions made and targets set as a result of political agreement about what should be done it is understood that they should not be open to contestation. This article argues, by contrast, that benchmarks are themselves highly political in a number of different ways. Firstly in that they have been constructed partly based upon certain political ideas about how best to govern; secondly in that, in related fashion, they reward certain practices and behaviours whilst ignoring others; and thirdly in that compliance with benchmarks has wide ranging politico-economic implications for compliance countries even in the instance that those countries were involved in the design of climate benchmarks.

This Special Issue is based upon the notion that different areas of benchmarking, according to governance domain, need to be understood on their own terms as well as collectively. In paying close attention to the specific ideas and assumptions that qualify and colour climate benchmarks this article moves beyond claims that global governance practices, like benchmarks, embody mainly market liberal orthodoxies. In order to better understand the precise nature of climate change benchmarks, this article first of all explores the somewhat mixed range of motivations behind the establishment of climate governance. It then reveals the complex mix of ideas and assumptions that have become embedded within benchmarks in order to explain *why* it is that certain politico-economic practices are seen and rewarded and others ignored by benchmarks. The intention here is not to argue for or against benchmarks but to reveal more about compromises reached and how these have limited the design and scope of targets and benchmarks.

Having revealed some of the character of climate benchmarks the next section explores what this actually means in practice. The approach taken here intends not just to assess whether benchmarks influence practice but to claim that the ways in which climate benchmarks have been designed are significant for domestic policy and politics but also, ultimately, for the legitimacy of climate governance. The method chosen here of making such assessments, and relating constructions to practices, is to study compliance at the domestic level in Germany and the UK, two developed and ostensibly climate aware countries. These countries are chosen partly because they

---

[1] Robert Garner, *Environmental Politics: the Age of Climate Change* (Basingstoke and New York: Palgrave Macmillan, 2011).

claim leadership in mitigation but also because they both score well and at similar levels in official EU benchmarking indexes – as such they are often held up as countries that pursue 'best practice' in climate governance.[2] Indeed, some climate indexes place the UK ahead of Germany, for example the high profile Climate Change Performance Index (CCPI) ranks the UK second only to Denmark whilst Germany languishes at number 19 (see Box 1).[3] This benchmarking position appears puzzling in the light of qualitative studies of climate policy and practices that rate Germany ahead of the UK.[4] By considering how climate benchmarks are constructed and how this inter-relates with compliance this article offers some explanation as to why such anomalies exist.

Global climate governance and embedded ideas

The approach to benchmarking taken here sits within broader constructivist IPE approaches to international governance and organisations (IOs), in that it explores how benchmarks are constructed rather than taking them as given.[5] Benchmarks are conceptualised here as embedded within but also as performing vital functions, of comparison and compliance, within processes of governing internationally. In this case climate benchmarks sit within governance attempts to set new and sustainable behavioural norms internationally. Benchmarks are further conceptualised as constructed objectivity, as reflecting existing knowledge, as power relations, and as weaving together the discursive and material worlds.[6] This latter is achieved through establishing quantitative benchmarking standards, often universally defined, and then encouraging compliance with those standards through measurement, comparison and, in some instances, naming and shaming.[7] Benchmarks as practice can have the ability to constitute a new common sense that narrows governance options down and normalises certain norms.[8] Arguably, however, in order for such processes to be successful there needs to be a degree of at least perceived legitimacy associated both with set benchmark standards and those actors that have set the standards.

---

[2] European Environment Agency (EEA), 'Report No. 10/2013: Trends and Projections in Europe 2013: Tracking Progress Towards Europe's Climate and Energy Targets Until 2020' (Copenhagen: European Environment Agency, 2013). See also International Energy Agency World Energy Outlook, 'Special Report 2013: Redrawing the Energy Climate Map' (Paris: International Energy Agency, 2013).

[3] J. Burck, F. Marten and C. Bals, *The Climate Change Performance Index: Results 2014* (Brussels: Climate Action Network Europe, 2014).

[4] See, for example, J. Mikler and N. Harrison, 'Varieties of capitalism and technological innovation for climate change mitigation', *New Political Economy*, 17:2 (2012), pp. 179–208; D. Toke and V. Laubner, 'Anglo-Saxon and German approaches to neoliberalism and environmental policy: the case of financing renewable energy', *Geoforum*, 38 (2007), pp. 677–87; J. Rosenow, N. Eyre, C. Rhode, and V. Buerger, 'Overcoming the upfront investment barrier – comparison of the German CO2 building rehabilitation programme and the British Green Deal', *Energy and Environment*, 24:1/3 (2012), pp. 127–36; C. Mitchell, D. Bauknecht, and P. M. Connor, 'Effectiveness through risk reduction: a comparison of the renewable obligation in England and Wales and the feed-in system in Germany', *Energy Policy*, 34 (2006), pp. 297–305.

[5] A. Broome and L. Seabrooke, 'Seeing like an international organisation', *New Political Economy*, 71:1 (2012), pp. 1–16; A. Broome, 'Constructivism in international political economy', in R. Palan (ed.), *Global Political Economy: Contemporary Theories* (London and New York: Routledge, 2013); H. K. Hansen and A. Mühlen-Schulte, 'The power of numbers in global governance', *Journal of International Relations and Development*, 15 (2012), pp. 455–65.

[6] A. Broome and J. Quirk, 'Governing the world at a distance: the practice of global benchmarking', *Review of International Studies*, 41:5 (2015), pp. 819–41.

[7] Ibid.

[8] L. Seabrooke, 'Pragmatic numbers: the IMF, financial reform, and policy learning in least likely environments', *Journal of International Relations and Development*, 15 (2012), p. 489.

As already observed elsewhere, however, there are differences between IOs – even between those operating within single policy areas.[9] A variety of modus operand, objectives and internal policy debates have been observed, as well as claims made that there is more than one Washington Consensus. Starting out from this position, this Special Issue as a whole disaggregates benchmarks by analysing practices separated out according to different transnational policy domains. A further method of disaggregating between benchmarks – suggested in the opening article of this Special Issue – is to group benchmarks according to typologies based on actors, who does the benchmarking, and the purpose or type of international practice that benchmarks are being used to actuate. Within this typology, the CPPI has been described as sitting within 'type IV' – conceptualised as being either explicitly or implicitly geared towards transnational advocacy in particular issue areas and used by a civil society organisation. There is certainly some merit in this description – climate benchmarks are indeed used as part of an overt attempt to make the world a better place by measuring countries against internationally agreed targets.

However climate benchmarks, in that they are so widely used by IOs to measure climate performance, also sit within 'type II' or 'benchmarking as international governance' and are less geared towards the promotion of individual state's national interests. It is therefore considered important to reveal here how benchmarks have been constructed, what practices at state level they reward and allow, and what implications this has in practice.[10] Given the degree to which benchmarks are used to establish and measure progress towards new norms by IOs and civil society it is important, for the success and legitimacy of climate governance, to assess whether benchmarks do, indeed, encourage behaviours consistent with mitigation and falling CO_2 levels. This is achieved here through disaggregating benchmarks in another way – by revealing and unpacking the knowledge structures that underpin them through an interrogation of EU climate governance norms.

This approach is based on the assumption that ideas are key variables capable of influencing governance choices – both at the domestic and international level – and that by being specific about ideas we can be more overt about *what* existing knowledge benchmarks are based on.[11] The emphasis here is on ideas as influential within processes of establishing new areas of governance given that climate benchmarks, and the targets upon which they are set, are part of the process of establishing new climate norms and of driving new and sustainable behaviours. Sociological institutionalists suggest that ideas play multiple roles during periods of, often domestic, policy change: alternative ideas can challenge existing orthodoxies by identifying weaknesses and problems and by suggesting alternative solutions to problems. But, importantly, embedded orthodoxies can also constrain and colour what types of new objectives and policy instruments are ultimately agreed and pursued.[12] By identifying these multiple roles for ideas we can understand that new governance norms are established through the push and pull between competing, but

---

[9] Broome and Seabrooke, 'Seeing like an IO', p. 8.

[10] Broome and Quirk, 'Governing the world at a distance'.

[11] On ideas and institutions see J. L. Campbell, 'Institutional analysis and the role of ideas in political economy', *Theory and Society*, 27:3 (1998), pp. 377–409; S. Berman, *The Social Democratic Moment: Ideas and Politics in the Making of Interwar Europe* (Harvard: Harvard University Press, 1998); M. Blyth, *Great Transformations: Economic Ideas and Institutional Change in the Twentieth Century* (New York: Cambridge University Press, 2002); C. Hay, 'Narrating crisis: the discursive construction of the winter of discontent', *Sociology*, 30:2 (1996), pp. 253–77.

[12] Blyth, *Great Transformations*; Hay, 'Narrating crisis…'.

not always compatible, sets of ideas revealing a more nuanced understanding of what knowledge international governance is based upon.[13]

Taking ideas and knowledge as influential is a common approach utilised in both international governance literatures but usually to describe path-dependencies. The forms of knowledge enacted through international governance processes are often described as generic. Such descriptions are not surprising given claims by some IOs that they are arbiters of knowledge about 'world's best practices' and 'good governance' and claims that countries performing well according to benchmarks also represent 'best practice'.[14] Legitimacy of norms and rules are inferred in the titles of 'best' and 'good' and it is also implied that countries following such practices will succeed in meeting targets. Such claims have led scholars to conceptualise some IOs as defining and enforcing governance according to 'Washington Consensus' norms, underpinned by liberal economic ideas about the role of the state in economic governance, economic efficiency, and fiscal responsibility.

Such claims are also common within some constructivist and IPE climate governance literatures – indeed it has been claimed that in order to gain acceptance internationally for climate action environmental concerns have been absorbed into the prevailing liberal economic order and associated power relations.[15] In this vein liberal economic ideas have been presented as erecting intellectual boundaries around the scope of policy imagination available to IOs in their everyday activities but also, importantly, around the design of benchmarks that seek to impact upon policy at the national level.[16] There is a small point of departure here in that climate governance is conceptualised as being informed by varied sets of ideas simultaneously, more in line with observations above about change taking place as a result of the push and pull between different sets of ideas. Climate governance is taken here not just as a case of decision-making during a process of change as bounded by one framework of ideas but as a coming together of different sets of ideas to motivate and inform the design of this emerging governance arena.[17] Multiple sets of ideas, environmental, scientific, and market liberal are embodied, enacted, and reified through climate benchmarks making this an emergent, highly political but complex area for analysis.

Such observations inform the methodology used here in that the article identifies and considers the influence of the different sets of ideas within processes of creating and utilising benchmarks at the international level. Environmental and scientific ideas are included to the extent that they have identified and outlined climate change as a problem that requires action thereby setting the new governance objective of reducing emissions.[18] Once accepted, however, this objective became coloured and constrained by orthodox knowledge within IOs about *how* to govern globally, that is, by setting targets, requiring countries to follow and assessing their progress using benchmarks. In a further stage ideas about the importance of economic growth to sustainability, about treating climate separately from environment, and about the need for energy

---

[13] C. Kuzemko, *The Energy Security Climate Nexus: Institutional Change in the UK and Beyond* (Basingstoke and New York, 2013), p. 180.

[14] Broome and Seabrooke, 'Seeing like an IO', p. 7.

[15] See, in particular, S. Bernstein, *The Compromise of Liberal Environmentalism* (New York: Columbia University Press 2001); but also H. Stevenson, 'India and international norms of climate governance: a constructivist analysis of normative congruence building', *Review of International Studies*, 37:3 (2011), pp. 997–1019 (p. 1001); and P. Newell, 'The elephant in the room: Capitalism and environmental change', *Global Environmental Change*, 21 (2011), pp. 4–6.

[16] Broome and Seabrooke, 'Seeing like an IO', p. 2.

[17] Kuzemko, *The Energy Security-Climate Nexus*, pp. 179–84.

[18] See also Stevenson, 'India and international norms of climate governance', p. 1001.

system transition are also shown to be embedded within climate norms. Understanding in more detail how ideas have structured benchmarks helps to reveal *what* practices are measured and encouraged at the international level as well as why. In many ways, however, this article suggests that what climate benchmarks do *not* see is as important as the behaviours that are encouraged.

In order to fully understand the practical significance of these constructions one further step is taken here by comparing processes of compliance with benchmarks in Germany and the UK. This is done in order to trace the substantive but variable effects of climate benchmarks *as constructed* – arguably a relative silence within existing benchmarking and climate governance literatures. Some climate governance scholars have suggested that the different domestic contexts and social, economic, and political processes involved in attempting to comply with climate benchmarks are consequential.[19] Rarely, however, is this claim explored in detail. It has also been claimed, within comparative climate scholarship, that different institutional arrangements have been better suited to building systems of sustainable change in response to climate targets.[20] Domestic ideas and institutions in this way colour how compliance takes place. What this article suggests is that such differences are not 'seen' by benchmarks because they are designed to simplify and quantify – the relative positions of Germany and the UK on EU climate benchmarks being a case in point. As such they are unable to recognise how domestic institutions affect a country's capacity and willingness to comply, or indeed the scale of political and energy system changes required. Furthermore, what activities are and are not measured by benchmarks becomes crucial to countries rankings and to which countries are held up as examples of good practice.

Benchmarks can and do, however, change and it is precisely for this reason that the politics of how compliance plays out in practice is important. IOs, governments, and NGOs have arguably been working from a far less established framework of ideas about what is right, what standards should be and how to measure them, given the somewhat unprecedented nature of climate mitigation as an area for governance. Climate agreements that have been reached, such as the EU's 20-20-20, have been hard won but are also are overtly subject to further renegotiation and potential fluidity as targets are all temporally fixed.[21] Much analysis of benchmarks assumes a one-way relationship inferring that policy is prior to compliance.[22] However, compliers with climate benchmarks are also important negotiators in setting new targets and this can be taken as one example of a 'recursive loop'.[23] In this way, how compliance happens, including resistances and contestations, really matters. International governance literatures, to the extent that they fail to consider these substantive and varied inter-relationships, stop short of fully realising the significance

---

[19] See also I. Bailey and H. Compston, 'Geography and the politics of climate policy', *Geography Compass*, 4:8 (2010), pp. 1097–14 (p. 1097); Garner, *Environmental Politics*, p. 5; P. Burnell, 'Democracy, democratization and climate change: Complex relationships', *Democratization*, 19:5 (2012), pp. 813–42 (pp. 815, 821).

[20] M. Lehtonen and F. Kern, 'Deliberative socio-technical transitions', in I. Scrase and G. MacKerron (eds), *Energy for the Future: A New Agenda* (Basingstoke: Palgrave Macmillan, 2009); Mikler and Harrison, 'Varieties of capitalism and technological innovation'; see also Garner, *Environmental Politics*, p. 4.

[21] In line with observations made by Stevenson, 'India and international norms of climate governance', p. 998.

[22] See O. J. Sending and J. H. Sande Lie, 'The limits of global authority: World Bank benchmarks in Ethiopia and Malawi', *Review of International Studies*, 41:5 (2015), pp. 993–1010.

[23] Broome and Quirk, 'Governing the world at a distance', p. 11.

of benchmarks for compliance countries but also interconnections between constructions, outcomes and further reconstructions of governance norms.

Climate governance and benchmarking as reformist

Global warming is one amongst a number of issues that have been raised and pursued by the international environmental movement, much of which has operated for many decades under a different philosophy to most other political perspectives.[24] Indeed, it is the environmental focus on inter-relationships between humans and nature that has helped us to conceive of climate change not as something inevitable but as an outcome of a range of human actions and as part of growing natural degradation. Such problems presented by humankind's actions have been conceptualised as global, given that as humans we live and operate in an interconnected fashion within the confines of the earth and its atmosphere – making global warming the ultimate global problem. What has further motivated a desire for global action is the emergence of a scientific consensus on anthropogenic climate change as well as lists of potential implications for humankind that have been associated with warming of above 2°C over preindustrial levels.[25] These notions have collectively come to form one of the bases of current international governance knowledge about climate change and the need for worldwide action.

It is important to note, however, that much environmental thought has also presented distinct normative challenges to global, as well as some domestic, governance orthodoxies. Environmental debates in the 1970s were informed by claims about natural limits to economic growth, whilst earlier literatures emphasised the links between capitalism, economic growth, fossil fuel use, and environmental degradation.[26] Indeed, models of growth since the industrial revolution have been based on ever-emergent new mechanisms and technologies that in turn were largely dependent on ever-greater energy use. From this perspective, modern capitalism is an integral part of the problem of climate change and solving it would for this reason require deep structural change – including a move away from common emphasis on economic growth.[27] This is what is referred to as a 'radical' environmental position.[28]

Despite growing degrees of élite political consensus around the notion that humankind is responsible for climate change and that action must be taken, the battle to establish an international climate framework was long and arduous and included many compromises from these original environmental and scientific debates.[29] Steven Bernstein explores in detail how environmental debates of the 1970s became imbued with existing power relations and liberal ideational orthodoxies emanating from

---

[24] Garner, *Environmental Politics*, ch. 1.

[25] IPCC (2007), in A. Giddens, *The Politics of Climate Change* (Cambridge and Malden, MA: Polity Press, 2009) p. 22; see also Stevenson, 'India and international norms of climate governance', pp. 997–8; Burnell, 'Democracy, democratization and climate change', p. 818. For example, one predication in Burnell's work is that global warming will result in 26 million 'climate displaced people' in the next twenty years.

[26] P. Newell and M. Paterson, *Climate Capitalism* (Cambridge: Cambridge University Press, 2010), pp. 12–13.

[27] J. Dryzek, D. Downes, C. Hunold, D. Schlosberg, and H. Hernes, *Green States and Social Movements: Environmentalism in the United States, United Kingdom, Germany and Norway* (Oxford: Oxford University Press, 2003).

[28] Garner, *Environmental Politics*, pp. 7–10

[29] Bernstein, *The Compromise of Liberal Environmentalism*, p. 32.

within international governance circles.[30] Global climate governance has through these processes been constructed through the co-option of only some environmental and climate science ideas whilst rejecting others. Indeed, amongst others it was those ideas that posed more of a challenge to Western liberal structures of growth and consumption that were sacrificed.[31] In these ways international climate governance, as currently structured, can be referred to as reflecting a 'reformist' position as it embeds the idea that it is possible to effectively incorporate climate protection within pre-existing political and economic structures of modern industrial society.[32] As explained in more detail below, from a radical position this approach is likely to result in failure to protect our environment and in governance measures that are too weak to make sufficient difference.

Through this compromised reformist approach international climate norms that have emerged, like other forms of international governance, are based upon setting international targets against which countries can be measured. Targets are understood here as specific, defined outcomes towards which (policy) efforts should be directed. It is assumed that the acceptance of legally binding targets will force change across a wide range of policy areas at the national level and ultimately produce changes in practices allowing for the rate of warming to slow and then be limited to 2°C.[33] Indeed, the notion that universal standards can and should drive other policy choices is deeply embedded within EU climate governance and the associated policy of mainstreaming. As EU Climate Commissioner, Connie Hedegaard, recently observed, '(c)limate solutions are all around – it's now on us to make them the norm'.[34]

In terms of specific targets, although new 2030 EU targets were agreed in October 2014, benchmarks have so far been based on the EU 20-20-20 package which includes a greenhouse gas (GHG) emissions reduction target of 20 per cent over 1990 levels. This emissions reduction commitment is above and beyond that of the Kyoto Protocol giving rise to claims that the EU is a leader in climate mitigation. However, the 20-20-20 package also includes other specific targets not included in Kyoto: to produce 20 per cent of EU energy consumption from renewable sources and to improve EU energy efficiency by 20 per cent both by 2020.[35] Here we can see the influence of ideas about inter-linkages between energy (as dominated by fossil fuels) and climate change – indeed these targets imply that energy systems must change in order to mitigate successfully.[36] These targets form the basis of various EU climate benchmarks, discussed in more detail below, and are accompanied by a range of preferred measures and instruments designed to enable countries to meet set targets

---

[30] Ibid.; See also P. Newell and M. Paterson, *Climate Capitalism: Global Warming and the Transformation of the Global Economy* (Cambridge: Cambridge University Press, 2010); Stevenson, 'India and international norms of climate governance', p. 999.

[31] H. Stevenson, 'Representing green radicalism: the limits of state-based representation in global climate governance', *Review of International Studies*, 40:1 (2014), p. 184.

[32] Garner, *Environmental Politics*, p. 8.

[33] M. Kneuer, 'Who is greener? Climate action and political regimes: Trade-offs for national and international actors', *Democratization*, 19:5 (2012), p. 865.

[34] Connie Hedegaard, *European Commission News* (2014), available at: {http://ec.europa.eu/commission_2010-2014/hedegaard/headlines/news/2013-12-20_01_en.htm} accessed 20 April 2015.

[35] New 2030 targets are: to reduce GHG emissions by at least 40 per cent below the 1990 level (binding at the EU level); to increase the share of renewables to at least 27 per cent (binding at the EU level); and to increase energy efficiency by at least 27 per cent (indicative only), available at: {http://ec.europa.eu/clima/policies/2030/index_en.htm} accessed 20 April 2015.

[36] It is estimated that up to 83 per cent of CO2 emissions in Annex 1 countries come from the energy sector – see IEA, *CO2 Emissions from Fuel Combustion: Highlights* (Paris: IEA, 2013).

such as the EU Emissions Trading Scheme (EU ETS) and the Energy Efficiency Directive.[37]

For some, given the emergence of scientific consensus on climate change and its causes, and associated studies about probable consequences, there should be no politics *of* climate change. Frustration is often expressed that possibilities already exist for climate mitigation to proceed but that politics, often defined separately and at the national level, actually stands in the way of achievements.[38] There are, furthermore, claims that climate change, given that it is so universal and important, should be taken out of 'partisan' political competition in order to give mitigation strategies the secure tenure needed in order to work.[39] This notion of climate change action being above politics coincides well with liberal institutionalist ideas about setting technical standards as 'objective' measures against which performance can be judged.[40] Climate action should, from these perspectives, be about technical transitions along measurable metrics towards lower emissions and a clean and sustainable future.[41] The approach taken in this article, however, denies the objectivity of climate targets and benchmarks but instead argues that their construction reflects politics, power relations, and complex and varied forms of knowledge about climate and energy.

Climate benchmarks and governance targets

There has been a proliferation of climate benchmarks over the past decade almost all of which are used for transnational advocacy purposes.[42] For example, the high profile CCPI, accumulated by Germanwatch, is used not only to assess and judge performance, against set standards, but also to enhance transparency and raise awareness of climate change via media channels and amongst political actors – with some naming and shaming also involved. Some attempt is made to highlight countries with best-practice policies – the inference being that less successful countries should follow their lead.[43] This makes it important that countries highlighted as 'good' are really on a long-term path to a more sustainable system. A large proportion of climate benchmarking is, however, undertaken by IOs. In fact all parties to EU, and Kyoto, agreements must be measured, as accurately as possible, against their commitments for the targets to have any practical effect.

EU climate benchmarks are highly quantitative and numbers oriented reflecting the targets upon which they are based – it requires a certain level of understanding of the science of climate change and of how different forms of energy function to comprehend them in any depth.[44] These benchmarks are put together annually to measure performance against GHG emissions reduction targets at the collective EU

---

[37] See DG Ener website, available at: {https://ec.europa.eu/energy/en/topics/energy-efficiency} accessed 20 April 2015.

[38] I. Bailey and H. Compston, 'Geography and the politics of climate change', *Geography Compass*, 4:8 (2010), pp. 1097–8.

[39] A. Giddens, *The Politics of Climate Change*, p. 189.

[40] See H. Compston and I. Bailey, 'Comparing Climate Policies: the Strong Climate Policy Index', Paper prepared for the PSA Annual Conference, Cardiff, March 2013.

[41] Garner, *Environmental Politics*, p. 6.

[42] See Broome and Quirk, 'Governing the world at a distance'.

[43] J. Burck, L. Hermwille, and C. Bals, 'The Climate Change Performance Index: Background and methodology', *Germanwatch* (November 2013), p. 4, available at: {https://germanwatch.org/en/download/8579.pdf} accessed 20 April 2014.

[44] See European Environment Agency, 'Trends and Projections in Europe: Tracking Progress Towards Europe's Climate and Energy Targets Until 2020', EEA Report No. 10/2013; Burck et al., *The Climate Change Performance Index*; Compston and Bailey, 'Comparing Climate Policies'.

level and against the individual targets that have been agreed for each member state. Aggregate EU, and individual country, renewable energy data is collected by a data collection agency, Eurostat, based within the European Commission.[45] These measures, along with efficiency data, are then combined in the annual European Environment Agency (EEA) report which serves as the principal EU climate benchmarking index. Interestingly when compiling benchmarks the EU often uses EEA data but they, in turn, rely heavily on information provided by member states.[46] The formal process of EU monitoring, reporting and verification (MRV) of GHG emissions takes place annually during the European Semester. Measuring against climate standards is, in addition, an emerging and inherently tricky business. With regard to measuring energy efficiency against targets different countries have varied criteria for assessing performance – not all of which are comparable.[47]

In these ways internationally agreed targets are central to climate governance but they also become the basis for climate benchmarks. In turn, benchmarks embody targets and become the principal method of measuring compliance success. Without measurement countries, and the EU, cannot be held accountable for missing targets. So intertwined are targets and benchmarks that in the language of EU climate governance binding targets are often referred to, in a quite inter-changeable manner, as benchmarks. The uneasy compromise between environmental ideas and liberal ideas about how to govern for global phenomena has so far resulted in an EU system designed to encourage behaviour change but using pre-existing governance methods.

EU climate benchmarks: What is seen and what is not

Technocratic, universal governance approaches have already attracted some criticism within climate change literatures.[48] This is partly because, like critiques of international governance practices in other policy domains, technical standards tend not to question the assumptions underlying them but also narrow down what human activity is seen.[49] This applies to EU benchmarks in that GHG emissions reduction, production of renewable energy, and improving efficiency have become what EU monitoring agents see and reward – but always with an important time lag and with little means of noting those countries that are working towards more progressive, long-term climate targets. Two other assumptions embedded in EU climate benchmarks, outlined below, further assist in delineating what is allowed but also what is not seen – assumptions that are shown here to both hide and actively approve behaviours that exacerbate climate change and environmental damage.

[45] DG Climate, 'Progress Towards the 2020 Targets: the European Semester' (2014) available at: {http://ec. europa.eu/clima/policies/g-gas/progress/index_en.htm} accessed 20 April 2015.

[46] European Environment Agency, 'Trends and Projections in Europe', p. 15; see also European Commission, 'Elements of the Union Greenhouse Gas Inventory System and the Quality Assurance and Control (QA/QC) Programme', Brussels 12 August 2013, SWD (2013) 308 final.

[47] J. Rosenow and R. Galvin, 'Evaluating the evaluations: Evidence from energy efficiency programmes in Germany and the UK', *Energy and Buildings*, 62 (2013), pp. 450–8

[48] V. Mathur, S. Afionis, J. Paavola, A. Dougill, and L. Stringer, 'Experiences of host communities with carbon market projects: Towards multi-level climate justice', *Climate Policy*, 14:1 (2014), pp. 42–62; P. Newell and A. Bumpus, 'The global political ecology of the clean development mechanism', *Global Environmental Politics*, 12:4 (2012), pp. 49–67.

[49] See J. Scott, *Seeing like a State: How Certain Schemes to Improve the Human Condition Have Failed* (New Haven, CT: Yale University Press, 1998), p. 76 in Broome and Seabrooke, 'Seeing like an IO', p. 7.

Economic growth is 'sustainable'

One important assumption built into climate benchmarks has to do with the accepted role of economic growth within Western capitalisms. Specifically, as part of the process of compromise discussed above, a decision was made to depart from the singular emphasis on environmental variables when defining what is sustainable to include economic growth as part of the definition.[50] This is evident within EU climate discourse to the extent that the framing of environmental issues rests on and includes a primary concern with economic growth.[51] The question of how climate change mitigation and economic growth relate to one another is by no means new in that it in many ways defines the split between climate reformists and radicals. Reformists and ecological modernisation theory suggest not only that economic growth is compatible with environmental protection but the two are, in fact, mutually reinforcing. In practice, reformist assumptions have led to 'win-win' rhetoric about climate and sustainable energy objectives being compatible with other EU institutions and a stress on the ability for countries to be simultaneously 'green and competitive'.[52] Such arguments stand in open contrast to environmental ideas about limits to growth, referenced above, that initially motivated climate governance.

It is hard to locate in detail the ways in which assumptions about the importance of economic growth have qualified climate benchmarks and how they are formulated – there are, for example, no specific economic growth metrics included and growth is as such not overtly rewarded. It is illustrative here, however, to consider briefly climate governance within the wider EU governance context. Climate targets have been made 'headline goals' of the Europe 2020 strategy, which is fundamentally about delivering growth, understood in terms of recovery from economic crisis. This growth, however, is qualified as needing to be 'smart, sustainable and inclusive'.[53] The qualification of growth as sustainable refers in particular to the move towards a low carbon economy and can been interpreted as part of the mainstreaming of new climate standards into other governance areas – in line with the assumption that growth can be green and with a win-win, reformist agenda.

The resultant position on growth is, however, slightly more complex in that the relationship between economic growth and climate standards as already suggested also works the other way – climate benchmarks actively incorporate an allowance for economic growth. Member states are judged against country emissions targets that are highly differentiated according to economic growth and development but also according to sectors of the economy. In this way benchmarks do recognise some difference between countries in terms of capacity to comply but principally according to metrics of developed or less developed according the GDP *per capita* and/or according to which industries are understood to be important to economic growth and therefore excluded. Countries like Bulgaria and Romania, for example, are benchmarked against targets that allow for them to continue to grow their emissions

[50] Bernstein, *The Compromise of Liberal Environmentalism*, pp. 70–1; A. Jordan, 'The governance of sustainable development: Taking stock and looking forwards', *Environment and Planning C: Government and Policy*, 26 (2008), pp. 17–33.

[51] Bernstein, *The Compromise of Liberal Environmentalism*, p. 53; Giddens, *The Politics of Climate Change*, pp. 193–4.

[52] J. Szarka, 'Climate challenges, ecological modernization, and technical forcing: Policy lessons from a comparative UK-EU Analysis', *Global Environmental Politics*, 12:2 (2012), pp. 87–109.

[53] European Commission, 'Europe 2020: Priorities' (2014), available at: {http://ec.europa.eu/europe2020/europe-2020-in-a-nutshell/priorities/index_en.htm} accessed 20 April 2015.

on the assumption that meeting tougher climate targets would necessarily involve high costs of implementing energy transitions which they could ill afford.[54] There are examples also of assumptions about the need to allow for growth benefiting developed countries. Some suggest that the choice of 1990 as the year against which GHG emissions reduction targets should be compared was made because it suited certain key counter-parties to the agreement – including Germany and the UK. Certainly, by March 2007 when the UK agreed to the 20-20-20 targets, it had already reduced its emissions heavily as a result of a switch from coal to gas during the 1990s and some de-industrialisation of its economy.[55] Measured in this way the UK meets emissions benchmarking standards and can be counted as a 'best practice' country.

Evidence of the way that commitment to growth has further qualified climate benchmarks can also be seen in the kind of targets agreed for energy efficiency and embedded within climate benchmarking standards. On the surface the commitment to a 20 per cent improvement in energy efficiency by 2020 looks impressive but it is, however, somewhat misleading. Firstly, unlike the other two targets the efficiency target is non-binding for individual countries. Secondly, it is measured not in absolute terms against a certain date, like GHG emissions, but as a reduction in energy consumption of 20 per cent below the *expected* energy consumption trend to 2020 and is, in these ways, a softer target.[56] There are a number of ways in which improving energy efficiency, and/or reduction in demand for energy, are considered to be related to economic growth. Given that growth in demand for energy has historically been considered a natural precondition of wider economic growth energy efficiency can infer, under some measurements, a reduction in growth.[57] For others, efficiency improvements may infer near and medium term costs – DG Ener estimates that the scale of investment required to meet EU 2020 efficiency targets is around €100 billion annually.[58] For many, this raises concerns about affordability and about the effect of these costs have on domestic households but also on companies' abilities to compete internationally. In the longer term, of course, greater energy efficiency is specifically supposed to allow for a decoupling between economic growth and energy demand growth.[59] Furthermore, for countries like the UK that have indigenous fossil fuel industries, there are also positive correlations between energy production, exports, tax receipts, and economic growth. This is one reason why fossil fuel production is still supported publically (often via tax breaks) in many countries. Although many organisations, like the International Energy Agency (IEA), have done much to quantify and discourage these practices, given negative implications for energy transition and for climate mitigation, such state support for fossil fuels is not measured directly by climate benchmarks and therefore is allowed.

Environmental campaigners, and some policymakers, have observed that politics gets in the way of meeting climate benchmarks at the national level as countries pursue growth over climate governance goals. It is observed, for example, that certain countries have in the past shied away from meeting targets or have diluted measures for growth reasons.[60] This suggests that meeting climate mitigation

---

[54] DG Climate, 'The 2020 Climate and Energy Package' (2014). This is related in turn to the principal of common but differentiated responsibilities and respective capabilities upon which the UNFCCC is based.
[55] Garner, *Environmental Politics*, p. 119.
[56] See DG Climate, 'The 2020 Climate and Energy Package'.
[57] Kuzemko, *The Energy Security-Climate Nexus*, p. 67; Newell and Paterson, *Climate Capitalism*, pp. 13–15.
[58] See DG Ener's webpage, available at: {https://ec.europa.eu/energy/node/1742} accessed 20 April 2015.
[59] Newell and Paterson, *Climate Capitalism*, p. 14.
[60] Compston and Bailey, 'Geography and the politics of climate change'.

objectives is understood in some countries as less important than meeting the objective of economic growth. This might not be a surprising outcome, however, given observations above that individual country climate targets have been specifically constructed with the intention of allowing for economic growth. In essence these choices are evidence of the softer reformist position having become embedded in climate benchmarks, thereby also avoiding and silencing other more radical interpretations of environmental problems and related solutions that recommend less economic growth and consumption.[61]

Climate as 'separate' from environment but as including energy

The second idea embedded in climate benchmarks, and what they measure, is that climate change can be considered as a separate area of governance from environment but as more inter-related with energy as a policy area. Evidence of the split between environment and climate can be seen in decisions to separate out the climate Directorate General (DG Clima) from the environmental DG but also in the design of other climate instruments. For example, by applying GHG emissions as the sole benchmarking standards CDM projects, that have had detrimental environmental impacts in the localities where they have been cited, can still be considered a success when measured from a distance.[62] Environmentally negative outcomes are simply not 'seen' as they are not pertinent to the measured standards but these impacts affect local communities in numerous ways.

For many, however, the most important side-effect of narrowly set climate benchmarks has been that nuclear power can, under such standards, be considered a 'clean' source of energy. It is not just that nuclear power is not seen but that it is considered by the EU to be a positive contribution to mitigating for climate change – contrary to the position taken in countries like Germany and Denmark. Clearly, from an environmental perspective, nuclear energy is potentially the most devastating form of electricity production available currently but it is also highly expensive.[63] The environmental implications of nuclear have become widely recognised during times of radioactive leaks for power plants but unanswered questions of where to store waste post decommissioning remain unanswered and may pose risks for future generations. As will be seen below, countries like the UK that produce nuclear are considered to be pursuing best practice whilst Germany, in its decision to completely phase it out, drop down benchmarks. In this way climate mitigation is, arguably, prioritised over environment.

Less critical attention has, however, been paid to the ways in which climate benchmarks, as separate from environment, then proceed to incorporate various

---

[61] For references to other claims that global climate governance silences alternative approaches to environment governance. see Bernstein, *The Compromise of Liberal Environmentalism*; B. Richardson, 'The governance of primary commodities: Biofuels certification in the European Union', in A. Payne and N. Phillips (eds), *Handbook of the International Political Economy of Governance* (Cheltenham: Edward Elgar, 2014); T. Wanner, 'The new "passive revolution" of the green economy and growth discourse: Maintaining the "sustainable development" of neoliberal capitalism', *New Political Economy*, 20:1 (2015), pp. 21–41.

[62] See Mathur et al., 'Experiences of host communities with carbon market projects', and Newell and Bumpus, 'The global political ecology of the clean development mechanism' for details of such projects and outcomes.

[63] A. Froggatt, C. Kuzemko, and E. Rouhaud, 'The energy security-climate nexus and the environment', in C. Mitchell, J. Watson, and J. Whiting (eds), *New Challenges in Energy Security: the UK in a Multipolar World* (Basingstoke and New York: Palgrave Macmillan, 2013).

assumptions about energy and its role in climate mitigation. Although the European Commission has separate DGs for energy and for climate change, as part of the creation of the new Energy Union there is now one commissioner for Climate Action and Energy. EU climate documents frame energy as a problem but also claim that renewables and efficiency, as core elements in a sustainable energy system transition, offer the primary route to reducing emissions.[64] What this article argues, however, is that what is not 'seen' here is the protracted and complex politics of driving energy policy, so long associated with fossil fuels, towards achieving climate benchmarks. Meeting even 'reformist' EU climate targets implies for most countries profound changes in energy governance, usage, and systems – albeit not as profound as changes implied within a radical agenda.

There are two aspects of this that are highlighted here. The first is that EU benchmarks, although they do make allowances for growth, tend not to see the *scale* of change required either socially, politically, or even in energy systems.[65] There are multiple separate but interconnected energy systems in place that have a bearing on how countries comply with targets embedded in benchmarks. Take, for example, the introduction of more renewable sources of energy into electricity systems. There are a host of primary and secondary institutions that support electricity systems each with their own implicit power relations, preferred technologies and abilities to influence governance decisions. Introducing more renewable electricity implies billions of euros of new investment just in new transmission systems, given the variability of renewable supply, let alone other parts of the electricity system. There are severe contestations and resistances against this type of change, which countries respond to according to domestic politics.[66]

The second aspect of what is not overtly 'seen' by benchmarks is the highly complex and messy politics of energy that need to be navigated in order to meet set standards. Significant political tensions between the EU and member states on energy policy have been noted elsewhere and although the EU has long worked to establish a unified energy policy there remain significant loopholes that countries can use to avoid compliance.[67] Article 194 of the Lisbon Treaty gives countries the sovereign right to choose, amongst other things, their energy mix.[68] This pits complying with renewable targets against not 'prejudicing' Member State's preferences for an energy mix that reflects their specific national circumstances. What is meant by this, and how it plays out in practice, is covered in more detail below but it is worth noting at this stage that both for the EU and for each member state attempts to meet renewable and efficiency targets relate in practice to attempts to meet other energy objectives.[69] These other energy policy objectives include the need to maintain

[64] DG Climate, 'Progress Towards the 2020 Targets: the European Semester' (2014), available at: {http://ec. europa.eu/clima/policies/g-gas/progress/index_en.htm} accessed 20 April 2015.

[65] For a discussion of the scale of change required, see J. Meadowcroft, 'Engaging with the politics of sustainability transitions', *Environmental Innovation and Societal Transitions*, 1 (2011), pp. 70–5 (pp. 73–4).

[66] See Broome and Quirk, 'Governing the world at a distance', p. 11.

[67] D. Buchan, 'Why Europe's energy and climate policies are coming apart', *Oxford Institute for Energy Studies*, SP 28 (July 2013); C. Kuzemko, 'Ideas, power and change: Explaining EU-Russia energy relations', *Journal of European Public Policy*, 21:1 (2014), pp. 58–75.

[68] European Commission, 'Consolidated version of the treaty on the functioning of the European Union', *Official Journal of the European Union* (30 March 2010).

[69] European Commission, 'A Roadmap for Moving to a Competitive Low Carbon Economy in 2040', Brussels COM(2011) 112/4 Provisional text.

energy security, to establish competitive markets, and sometimes also to address serious energy poverty issues.[70]

In practice, therefore, although the EU and member states do seek to meet climate benchmarks there are other important but differential hierarchies of what is considered politically important, plausible, and possible. Assumptions are built into benchmarks about energy efficiency and renewable energy being solutions to *both* climate change and energy security problems but again this win-win rhetoric masks complex interactions between energy policy objectives in practice.[71] What then also becomes important is the wider context within which climate governance sits and analysis of how governance areas interact with one another, including what priorities are chosen, within processes of compliance with benchmarks.

Member state compliance

One of the puzzles outlined in the introduction to this article is that Germany ranks similarly to, or in some instances much lower than, the UK in climate benchmarks and this section explains in more detail why such rankings seem questionable to many climate scholars (see Table 1 for rankings).

	CPPI 2014	EEA 2013 Efficiency	EEA 2013 Renewables	EEA 2013 Emissions
Germany	19th	Good Progress	Indicative targets met	Mixed Progress
UK	2nd (5th)	Some Progress	Indicative targets not met	Good Progress

Note: Germanwatch's CPPI does not assign positions 1 to 3 to any countries – they consider that no countries are doing enough to prevent dangerous climate change. As such the UK is at position 5 but it comes second relative to all other countries covered.

Table 1. *German and UK Performance in CPPI and EEA benchmarks*

The EEA's latest climate benchmark, in ranking Germany and the UK highly, infers that these are countries whose example should be followed. If these countries are held to be model reformers a critical exploration of the politics of compliance is indeed a useful exercise – especially given that climate mitigation is a relatively untested area of governance where much learning will need to take place. The importance of national compliance has been emphasised elsewhere in that international targets are understood to have little effect on emissions until policies are introduced at the national level.[72] The politics of compliance is also taken here as important as EU member states, often as part of negotiating blocks, have much to say about what should and should not be measured in future based upon national experiences with complying.

Given also that climate targets and benchmarks are intended to set a new direction against which countries can be measured it is worth noting that, even against reformist targets, EU benchmarks do reveal some slippage. Measured compliance with EU GHG emissions reduction targets show that although the aggregate EU emissions reduction target for 2020 is likely to be met, 13 member states may not meet

---

[70] The EU's formal energy policy objectives are: competition, energy security, and climate mitigation.
[71] Froggatt et al., 'The energy security-climate nexus and the environment'.
[72] N. Carter, 'Climate change and the politics of the global environment', in M. Beeson and N. Bisley (eds), *Issues in Twenty-first Century World Politics* (Basingstoke and New York: Palgrave Macmillan, 2010), p. 57; Compston and Bailey, 'Comparing Climate Policies', p. 3.

their individual emissions targets.[73] In this way the performance of some countries subsidises other non-performers. Aggregate EU and individual member state compliance with renewable energy targets is yet more mixed, and even worse is energy efficiency compliance, albeit efficiency targets are non-binding.[74] Failure to meet these 'weak' targets may have significance given the EU's 'leadership' position in climate negotiations and commitment to climate mitigation.

German compliance: the 'Green' reformer

EEA climate benchmarks tell us that Germany, on the latest measurable data, is making good progress towards meeting EU targets – especially with regard to renewable energy.[75] It has for some time been qualified as one of the best EU countries in compliance terms – although it has slipped recently on emissions measurements for reasons to be explained below.[76] As already suggested, however, what is not seen by benchmarks is the amount of work, in terms of policies and institution building, that has already gone into meeting these targets. Germany, unlike most other countries in the world including the UK, is already in 'phase II', the Energiewende, of its energy transition.[77] The Energiewende is specifically designed to facilitate a sustainable transition by 2050: it contains a GHG emissions reduction target of 80 to 95 per cent and a reduction in primary energy use of 50 per cent. These targets are notable in that they are both measured against 2008 rather than the 'easier' date of 1990 and therefore infer a high degree of change. It also includes one of the most ambitious renewable energy targets of 60 per cent of final consumption.[78] What this shows is long-term commitment above and beyond EU benchmarks – something that is neither seen nor rewarded because benchmarks only measure against EU agreed targets and as such only recognise, on an annual basis, success in meeting these. Ambitious climate targets can be understood as part of the broader German history of having tougher environmental targets than the EU.[79]

What compliance emphasises is that having targets is really only a starting point and that learning how to meet them is the important part. Here Germany is widely held up, in more qualitative climate analyses, as a clear leader. For example, Germany's feed-in-tariff (FiT), the principal support mechanism for renewable energy, has been highly successful, not least in that it has offered a risk-free return for new renewable production. Germany has also supported renewable production by ruling that transmission systems must give priority access to renewables over other sources of electricity.[80] These conditions have made it possible for a high degree of

[73] DG Climate, 'Progress Towards the 2020 Targets: the European Semester' (2014), available at: {http://ec.europa.eu/clima/policies/g-gas/progress/index_en.htm} accessed 20 April 2015; European Environment Agency, 'Trends and Projections in Europe'.

[74] European Environment Agency, 'Trends and Projections in Europe', p. 124.

[75] Ibid., various pages.

[76] Ibid., p. 111.

[77] Phase one being the 'Energy Concept' and Renewable Energy Act of 2000, see G. Fuchs, N. Hinderer, G. Kungl, and M. Neukirch, 'Adaptive Capacities, Path Creation and Variants of Sectoral Change: the Case of the Transformation of the German Energy Supply System', SOI Discussion Paper 2012-02 (2012), p. 19.

[78] Agora Energiewende, '12 Insights on Germany's Energiewende', Agora Energiewende Discussion Paper 010/03-1-2013/EN (2013), p. 1.

[79] V. Schmidt, *Democracy in Europe* (Oxford: Oxford University Press, 2006), p. 145.

[80] D. Jacobs, 'The German Energiewende: History, targets, policies and challenges', *Renewable Energy Law and Policy*, 3:4 (2012), pp. 223–34.

small-scale but also widely distributed production to be developed in Germany approximating more closely the environmental ideal of community and/or local energy production. Although benchmarks show that Germany is meeting its renewable targets this more sustainable aspect of how renewable energy is produced is neither measured nor rewarded by EU climate benchmarks. Ability to comply here is also important in that Germany is a now strong supporter of extending binding renewable energy targets at the EU integrated and member country level to 2030 and 2050.[81]

In attempting to explain why Germany has been more successful in supporting renewable energy comparative analyses suggest that as a 'co-ordinated market economy' Germany has already had in place the kinds of institutions that allow for and support this kind of long run but difficult and constantly renegotiated change.[82] Not only has there been considerable public investment in research and development of new systems they have also had better coordinated processes for deliberation and negotiation in support of new, but sometimes contentious, policies. This is partly also because Germany has recognised the need for specific institutions to enable a sustainable transition. The well-capitalised German development bank, the KfW, has been directed to lend to sustainability projects at very low interest rates and then reinvest proceeds back into more sustainability projects.[83] Furthermore Germany has a more deeply embedded municipal movement that has supported the many local energy projects that have developed across Germany, albeit more specific to some Länder than others. Some also argue that the costs of energy transition are more fairly distributed to the extent that those that are impacted are better supported by the German welfare system with positive implications for energy poverty.[84] This has all been further underpinned by the fact that all political parties, as well as large sections of the German electorate, support climate change mitigation, energy system transition, as well as other environmental issues.[85]

Beyond this institutional terrain a close analysis of the politics of compliance reveals deep complications. This is where we pick up again on some of the qualifications to climate benchmarks outlined above. One assumption built into climate benchmarks is that climate targets are *the* drivers for cleaner energy policy and systems. Historically, however, improvements in energy sustainability have been a means of responding not to climate but to energy security fears. It has also been suggested that much of what has been achieved in Germany, and other countries like Sweden, has been related to their lack of indigenous energy supply and the fact that they have historically been heavy fossil fuel importers. The argument runs that the initial impetus for some measures was to reduce dependence, and thereby increase supply security, in response to crises such as the 1970s oil shocks.[86] In not recognising

---

[81] See D. Buchan and M. Keay, 'The EU's Energy and Climate Goals for 2030: Under-Ambitious and Over-Bearing', Oxford Energy Comment (January 2014).

[82] Mikler and Harrison, 'Varieties of capitalism and technological innovation'; M. Lockwood, C. Kuzemko, C. Mitchell, and R. Hoggett, 'Theorising Governance and Innovation in Sustainable Energy Transitions', EPG Working Paper No. 1304 (2013), available at: {http://projects.exeter.ac.uk/igov/working-paper-theorising-governance-and-innovation-in-sustainable-energy-transitions/} accessed 20 April 2015.

[83] For an assessment of the KfW Bank's role in facilitating system change, see M. Schröder, P. Ekins, A. Power, M. Zulauf, and R. Lowe, 'The KfW Experience in the Reduction of Energy Use in and CO2 Emissions From Buildings: Operation, Impacts and Lessons for the UK', UCL Working Paper (November 2011).

[84] M. Lockwood, 'The Political Dynamics of Green Transformations', Exeter Energy Policy Group Working Paper No. 1404 (April 2014).

[85] Giddens, *The Politics of Climate Change*, p. 76.

[86] Ibid., p. 37.

different drivers for change climate benchmarks see only a small portion of relevant decision-making around energy policy.

A second embedded assumption that has an impact on how Germany is judged is that climate can be separated from environment. There has been a long-standing debate in Germany about nuclear power – in the early 2000s it was agreed that nuclear power, despite environmental objections, could act as a short-term, low carbon bridge to a future where energy consumption would be lower and renewables the primary source.[87] In 2011, it was decided that nuclear would be phased out by 2022 – a controversial as well as momentous decision that reflects a more environmentally informed standpoint on energy production.[88] Nuclear power in 2011, however, still accounted for 17.7 per cent of Germany's electricity supply and this has therefore inferred profound short and medium term changes to energy policy and systems – not least in that its options for electricity production have narrowed. In the short-term, to accommodate for nuclear power shortfalls, Germany has been producing more electricity from coal and for this reason fell foul of emissions measurements.[89] By extension, however, this also places the need to develop renewable energy quickly absolutely central within German climate and energy policy and infers economic and political costs – the ramifications of which are currently very high profile. EU climate benchmarks, because they do not measure environmental impacts, only see the withdrawal from nuclear as negative decision in that it has implications for near-term emissions. The EU, for its part, is critical of Germany's stance on nuclear and the degree to which Germany subsidises renewable energy production.[90] In this benchmarks are both near-sighted, in that they do not account for longer-term commitments, but also because in only recognising emissions, but not other environmental impacts of energy, they allow and even reward potentially damaging practices thereby storing up future problems. This shows the impact, in practice, of prioritising climate over environment in this way.

Despite Germany's position as one of the most progressive in terms of action to mitigate for climate change there has always been deep seated opposition but the German Parliament's resolution in addressing this is not rewarded. The 'Big 4' gas and electricity companies, the coal sector, and some heavy industry companies have mounted, and continue to mount, sustained attacks on green energy policy decisions.[91] As a mark of German parliamentary commitment to sustainable transition and renewables, MPs have actively defended renewable policies against the 'Big4'.[92]

Those that oppose the Energiewende today use the economic costs associated with rapid energy transition to argue against further support. Opponents of renewables place the blame for high-energy prices in Germany on the cost of support mechanisms which are passed on to customers.[93] One report argues that Germany has lost €15 billion annually in trade due to premium energy prices

[87] D. Buchan, 'The Energiewende – Germany's gamble', *Oxford Institute for Energy Studies*, SP 26 (June 2012), p. 3; Fuchs et al., 'Adaptive Capacities, Path Creation and Variants of Sectoral Change', p. 19.

[88] As does the decision to limit bio-energy to less than 10 per cent of power generation in the long term. See Agora, '12 Insights on Germany's Energiewende', p. 5.

[89] European Environment Agency, 'Trends and Projections in Europe'.

[90] Buchan, 'Why Europe's energy and climate policies are coming apart', p. 2.

[91] Fuchs et al., 'Adaptive Capacities, Path Creation and Variants of Sectoral Change', p. 21.

[92] S. Jacobsson and V. Lauber, 'The politics and policy of energy system transformation – explaining the German diffusion of renewable energy technology', *Energy Policy*, 24 (2006), pp. 256–76 (pp. 265–6).

[93] Fuchs et al., 'Adaptive Capacities, Path Creation and Variants of Sectoral Change', p. 19.

faced by industry.[94] This is despite the fact that, in terms of heavy industry, concessions have been made.[95] The economic costs of transition, as well as other system costs are not passed on to car, chemicals, and steel industries in the same way as they are to other, especially residential, electricity consumers.[96] In these ways we see some relative priorities come to light: energy transition is not prioritised over heavy industry, historically such a large part of German GDP, on the grounds of economic growth and international competitiveness arguments. Allowing for growth in this way is, however, in line with assumptions about sustainability that are built into benchmarks and as such countries are free to follow this path.

What this shows is that even pursuing a reformist climate agenda in a relatively well committed country is, in practice, highly political as well as significant for other governance areas. The future of the Energiewende at the moment appears viable, but new opposition has started to mount as economic costs, as currently distributed, impact upon domestic over industrial users. Most recent changes have been to pull back on the generous FiT system – although priority access for renewables remains in place – partly also in response to EU pressures to reduce state support levels. Their future ability to meet emissions targets will hang on how successful they are in balancing nuclear phase out with renewable development and demand reduction. In this they are facing unprecedented but deeply difficult decisions – their experiences may provide useful learning for countries that are considering a similar route. According to climate benchmarks, all this commitment of political capital in the face of opposition is simply not relevant thereby making energy system transition in pursuit of climate mitigation appear technical and straightforward.

UK compliance: Reluctant reformer meets status quo

As we have seen, the UK is assigned a position in official EEA benchmarks similar to Germany and, as such, it is also often held up as a leader in acting to mitigate for climate change. In assigning the UK a position well ahead of Germany the CCPI index suggests that it is much better at climate mitigation. By contrast, however, and as already suggested in the introduction to this article, comparative analyses of climate governance are more critical of the UK's approach to climate mitigation. EU climate benchmarks arguably see even less of the politics of UK compliance, in particular the internal contestations even of reformist EU targets. For example, although the UK has historically performed well against emissions reduction targets it is argued that these achievements have mainly been based on the large-scale switch from coal to gas that took place in the 1990s as well as on the recession that started in 2008, but less on proactive policymaking.[97] In fact, estimated figures for 2012 show a rise in UK GHG emissions due to a switch back to coal driven by the relatively low coal versus gas prices but benchmarks, given the time lag, only see such deterioration years later.[98] In addition not only is UK's 2020 renewable energy target of 15 per cent lower than Germany's 18 per cent it is also, at 5.6 per cent in 2013, still a long way off

---

[94] J. Vasagar, 'Germany Told of Billions Lost to Trade Due to Energy Policy', *Financial Times* (26 February 2014).
[95] Giddens, *The Politics of Climate Change*, p. 195.
[96] Buchan, 'Why Europe's energy and climate policies are coming apart', p. 4.
[97] D. Helm, *Energy, the State and the Market: British Energy Policy Since 1979* (Oxford: Oxford University Press, 2003).
[98] European Environment Agency, 'Trends and Projections in Europe', p. 33

reaching it.[99] Climate benchmarks in this way reward the UK for lowering emissions and making progress towards renewable targets but fail to make any relative judgement in favour of Germany's progress in meeting tougher renewable targets than the UK.

Another reason for the relative favouring of the UK within EU benchmarks, as currently constructed, might be that the UK's orientation toward target setting is more in line with the EU. The Labour government, under Gordon Brown, adopted a legally binding emissions reduction target of 80 per cent by 2050 as part of the Climate Change Act and this has been held up as best practice. Although it should be noted that the 80 per cent figure is measured against 1990 levels and is therefore not a commitment comparable to that being made by Germany. In order to meet these objectives new institutions have recently been established, such as the Department for Energy and Climate Change (DECC) and the Green Bank and a plethora of new white papers, strategies, bills, and acts of parliament has also ensued. Broadly speaking, however, the UK is still in phase one of a weak reformist programme on climate mitigation. A return to climate and energy policy literatures suggests that the UK's new climate targets and recent governance changes have been less successful at driving actual changes in the energy system.[100] This infers that target setting and energy system change do not necessarily go hand in hand but also that there should be greater questions posed about the UK's status as 'best practice' in climate mitigation.

One prominent explanation given for relatively disappointing energy system change has been that the UK has placed a comparatively greater regulatory and policy emphasis on other energy policy goals, such as liberalising and privatising markets, and on the role of markets in delivering sustainable transition and energy services more broadly.[101] In the 1990s as privatisation progressed responsibility for energy was passed to private enterprise and the UK closed energy departments down thereby arguably limiting civil service and wider government energy sector knowledge. Privatisation for some has also fostered an oligopoly of private gas and electricity companies, the 'Big 6', which have been reluctant to change and which have had a high degree of influence over policy and regulation.[102]

This relative stance versus stated political intentions to transition to a sustainable energy system stands in strict contrast to German parliamentary standing down of large utilities. One example of corporate influence over UK policymaking has been the recent government decision to pull back on energy efficiency obligations that had been placed on energy companies in response to popular sensitivity to rising prices. This was a consummate piece of deflection by the Big 6 given that, at the time, many market and political commentators were pointing the finger at energy corporates as culpable for the price increases not climate policies. The fact that the UK government gave way so easily on efficiency policy prioritises suggests that powerful industries and popular opinion about the price of energy are considered more important than climate mitigation.

---

[99] Department of Energy & Climate Change, 'Renewable sources of energy: Chapter 6, Digest of United Kingdom Energy Statistics (DUKES)' (London: DECC, 2015).

[100] See C. Mitchell, *The Political Economy of Sustainable Energy* (Basingstoke and New York: Palgrave Macmillan, 2008) for an in-depth analysis of UK sustainable energy policy and disappointing changes in actual practices in the energy system.

[101] See, for example, Mitchell, *The Political Economy of Sustainable Energy*; Giddens, *The Politics of Climate Change*; see also various chapters in I. Scrase and G. MacKerron (eds), *Energy for the Future: A New Agenda* (Basingstoke and New York: Palgrave Macmillan, 2009).

[102] Lockwood et al., 'Theorising Governance and Innovation in Sustainable Energy Transitions'.

This analysis of the politics of UK compliance highlights, again, the importance of notions of energy security as drivers for change but also as capable of influencing choices between low carbon forms of energy. For a period, mainly in the 1990s, the UK was not only able to cover its domestic needs for oil and gas but was also able to export fossil fuels and this reduced its exposure to security of supply issues. It wasn't until the mid 2000s, when becoming a fossil fuel importer again coincided with other international energy events, that the UK became highly aware of such concerns and made energy security, alongside climate change, a priority energy policy objective. The resulting political emphasis on 'home-grown' energy, as capable of improving security, has been utilised by the nuclear industry to push for government support for new UK nuclear power stations. Contrary to observations about the UK's commitment to markets, the Conservative-Liberal Democrat Government have agreed to long-term, fixed prices, at double the market rate, for nuclear electricity.[103] This suggests a prioritising of energy security and nuclear over renewables given that new nuclear will provide for the low carbon element that would otherwise have had to come from renewables. EU benchmarks not only allow for nuclear, given its low carbon credentials, but do not see the ways in which commitment to nuclear can reduce commitment to renewables.

When considering a country's ability to comply existing energy systems and infrastructures are also important.[104] The UK's historic ability to earn export revenues from fossil fuel exports, embedded infrastructures and associated sunk costs impact upon existing power relations in energy and give succour to arguments against energy transition. The UK, furthermore, supports the domestic fossil fuel industry with billions of pounds of subsidies and tax breaks.[105] There has been a recent decision to design the most generous taxation system on offer for shale gas to facilitate indigenous exploration and development as well as to establish a new Oil and Gas Agency to facilitate maximum possible extraction of oil and gas in the UK.[106] These are all directly framed not only as being necessary in order to secure energy security but also to secure economic growth, or more precisely recovery. Such investment in fossil fuels to support economic growth is allowable under current constructions of EU benchmarks and standards as well as in line with the UK's own definitions of sustainability.[107] As such, decisions affecting compliance with climate targets are not just about an over-reliance on markets but also about inter-actions between energy policy objectives of emissions reduction, supply security and the objective of economic growth. Although the European Commission has had much to say about UK nuclear subsidies,[108] the EEA's benchmarks do not mark the UK down for actively supporting fossil fuel expansion that is detrimental to climate change mitigation.

[103] See the Gov.UK website, 'Initial Agreement Reached on New Nuclear Power Station at Hinkley', Press Release (21 October 2013), available at: {https://www.gov.uk/government/news/initial-agreement-reached-on-new-nuclear-power-station-at-hinkley}.

[104] See G. Unruh, 'Understanding carbon lock-in', *Energy Policy*, 28 (2000), pp. 817–30 for a discussion of carbon path-dependencies or 'lock-in'.

[105] House of Commons Environmental Committee, *Energy Subsidies: HC61* (London: the Stationary Office 2013), p. 3. The UK is, however, uncertain in reality about how much it spends on subsidising energy including nuclear, fossil fuels, and renewables – or even how this should be measured. The investigation into subsidies is still ongoing.

[106] BBC News, 'Aberdeen to be Home for New Oil and Gas Agency', available at: {http://www.bbc.co.uk/news/uk-scotland-north-east-orkney-shetland-27809836} accessed 20 April 2015.

[107] Kuzemko, *The Energy Security-Climate Nexus*, p. 98.

[108] P. Johnstone, 'European Commission Critique of UK Nuclear Strategy', *Sussex Energy Group Blog* (11 February 2014), available at: {http://sussexnrggrp.wordpress.com/2014/02/11/the-potential-for-a-hinkley-shaped-hole-in-uk-energy-infrastructure/} accessed 20 April 2015.

This support is simply not seen and, in this way, questions benchmarks in both efficacy and credibility terms.

In terms of choices regarding energy mix, that is what sources should be prioritised, this is far from a coherent transition agenda. By contrast by rejecting nuclear and emphasising renewables Germany has made choices about technology pathways and this is important in that it then allows other decisions to be made about what associated energy system changes will be needed. The UK has not made firm decisions about energy sources in that everything is still on the table but it also does not have any renewable energy targets beyond EU 2020, certainly not binding ones. By not choosing variable renewable over steady-state, coal, gas and nuclear, electricity supply the UK has to maintain transmission and distribution systems that cater to both – arguably a more expensive option in the long run. This lack of commitment to renewable energy has underpinned recent attempts to convince the European Commission that no binding renewable targets should be placed on member states as part of the 2030 framework thereby seeking to alter the metrics against which climate benchmarks measure compliance. In these ways, although the UK is judged as similar to or better than Germany, it should be described as more *status quo* even than reformist at the moment and should not, perhaps, be held up as a model of climate mitigation performance.

Conclusion

This analysis has, by adopting a constructivist IPE approach, been able to reveal the many ways in which climate benchmarks, as currently constructed, have implications for which countries are held up as leaders, how compliance is undertaken, as well as complex inter-linkages with other policy areas. By narrowing down what is measured to numbers reflecting certain, soft targets Germany becomes, in one benchmark, less of a model climate reformer than the UK – encouraging other countries to follow UK practices. This is problematic to the extent that other, more qualitative, analyses claim that UK climate and energy governance is in many respects less effective than Germany's. There are issues here of temporality, to the extent that each benchmark assessment exercise focuses only on one year, and of scale of ambition, in that targets accepted tend to delineate the boundaries of climate ambition expected of each country – we will return to these issues below. Largely this is a case of benchmarks reflecting an acceptance of what is held to be achievable based on today's politico-economic models while glossing over more thorny but none-the-less equally important issues for climate mitigation and sidelining radical alternatives. This is not so surprising given that, once recognised as a global issue that must be addressed, it has fallen to pre-existing international institutions to take action but may not bode well for long-term climate mitigation.

The approach taken here has departed from some constructivist IPE approaches to the extent that climate governance and benchmarks have been conceptualised as informed simultaneously by multiple, varied ideas. In particular there has been an emphasis on revealing the internal contradictions that occur between intentions to drive sustainable behaviour change and those behaviours benchmarks actively reward and those that they ignore. A number of specific issues have been highlighted – one is that climate benchmarks infer an active but sadly limited inter-relationship between policy areas. The only explicitly recognised link between policy areas is that between

energy and climate policy but the way in which energy is included in climate benchmarks ignores other aspects of energy policy that are equally important to the success of climate mitigation strategies. Countries are understood to be following best practice if they successfully pursue sustainable energy policies but they are not directly discouraged from supporting fossil fuels given that such energy policies are simply not seen. Some links between climate and other policy areas are broken, for instance with environmental policy allowing for climate choices that are in effect environmentally damaging, whilst other links are simply inferred such as those between economic growth objectives and meeting climate targets.

A second point of departure from climate governance literatures has been to consider compliance in two developed countries in order to make explicit how different domestic and institutional contexts inter-react with international standards and influence different compliance paths. This analysis has revealed nuanced differences in German and UK climate governance – not least Germany's greater commitment to long-term energy system transition as opposed to the UK's tendency to attempt to comply rather than lead but also to question even existing, soft targets. Indeed reifying economic growth benchmarks in essence allows countries, like the UK, to object to climate mitigation agendas on precisely those grounds. For example there has been heavy objection from the UK over renewable energy's place within EU 2030 targets as well as over the inclusion of efficiency targets.[109] By conceptualising EU climate governance as fluid and as involving two-way relationships with member states we can therefore understand that current benchmarking constructions, and the contestations they engender and allow, become part of their future construction.

For those from a radical environmental standpoint missed targets and disagreement over the near-term future might be all too predictable but that still leaves the international community with the ever more desperate need to meet the 2°C limit and to further institutionalise behavioural change. The suggestion here is to more overtly recognise precise ways, outlined above, in which benchmarks are currently limited in their ability to recognise effective climate mitigation strategies. Once such limits are made more overt there could be two basic options for proceeding: one would be to take benchmarks with a large pinch of salt and not use them to quantify 'best practice' but this would necessitate greater structural reforms given their central role in measuring compliance with binding EU targets. In the absence of such reforms the second option would be to improve their construction. Benchmarks might, for example, be altered to include *new* standards against which each country's climate performance can be measured, not least standards that can reflect how fossil fuel production and use is developing in each country. New standards that measure the degree to which countries are managing to decouple economic growth from energy demand growth, and or measure each country's longer-term progress and future commitment to climate mitigation might also be considered. These suggestions are based on the analysis undertaken here, but should by no means be taken as a complete list.

Devising these would require further knowledge building about the complex detail of how benchmarks are constructed and how different nations interpret and comply with climate benchmarks in practice. From a theoretical viewpoint taking an approach that better reveals embedded ideas and interrogates how they relate to

---

[109] See O. Geden and S. Fischer, 'Moving Targets: Negotiations on the EU's Energy and Climate Policy Objectives for the Post-2020 Period and Implications for the German Energy Transition', SWP Research Paper, RP3 (March 2014).

policy constructions has arguably allowed for a more honest and realistic assessment of climate governance. More such analysis might be useful in further revealing complex constructions of climate governance and varieties of compliance and make more overt what political choices are being made and the hierarchies between economic, energy, environmental, and climate objectives these choices infer. Such knowledge, if it were to be made more readily available, would allow voting publics to know more about what compromises governments are making (and need to make) in their name. All these observations imply that what is needed here is a reflexive political learning process, instead of assuming that answers are already known and can be technically measured, especially given the unprecedented nature of governing for climate mitigation and complex system transition.

The limits of global authority: World Bank benchmarks in Ethiopia and Malawi

OLE JACOB SENDING AND JON HARALD SANDE LIE

Abstract. Global benchmarks (re)shape political conversations and institutionalise authoritative languages. It does not necessarily follow, however, that benchmarks can exert a lasting or significant influence over policies and behaviour of benchmarked actors. We analyse how the World Bank uses benchmarks to manage its relations with both donors and recipient governments. We analyse the role of the World Bank's Country Policy and Institutional Assessment (CPIA), both at headquarters and in relation to the recent history of two countries in Africa: Ethiopia and Malawi. We find that the CPIA is not – and contrary to what one would expect from the CPIA's nominal function and the literature on benchmarks – a very important tool for signalling incentives and allocating funds, or shaping the policy dialogue or the World Bank's strategy in these two countries. Rather, the CPIA is used highly selectively as one factor among many in the negotiations between World Bank staff and governments. We conclude that the CPIA helps establish the World Bank as an actor that embodies global authority on development issues, including with donors, but that there is a tension between such global authority on the one hand, and concrete authority to shape policy in domestic contexts, on the other.

Ole Jacob Sending is Director of Research at the Norwegian Institute of International Affairs (NUPI). His research focuses on the role of international and non-governmental organisations in global governance. His work has appeared, *inter alia*, in journals such as *European Journal of International Relations*, *International Political Sociology*, *International Studies Quarterly*, and *International Theory*. Recent works include *The Politics of Expertise: Competing for Authority in Global Governance* (University of Michigan Press, 2015), and *Diplomacy and the Making of World Politics* (co-edited with V. Pouliot and I. B. Neumann) (Cambridge University Press, 2015).

Jon Harald Sande Lie is a social anthropologist and Senior Research Fellow at the Norwegian Institute of International Affairs (NUPI). His research focuses on aid relations, the postcolonial, development state in Africa, and the practice of humanitarian principles. His research has appeared in, amongst others, *Third World Quarterly*, *Millennium*, *Social Analysis*, and *Africa Security*, and his monograph *Developmentality: An Ethnography of the World Bank-Uganda Partnership* is out on Berghahn Books (September 2015).

Introduction

Benchmarking is a governance instrument that – among other things – establishes a shared reference point among otherwise distinct and distanced actors. The use of benchmarks and indicators to monitor and govern from afar has proliferated over the last two decades, and the international development apparatus is a striking case in point. Since the early 2000s, donor institutions have increasingly relied on benchmarks in their engagement with client countries, partly because of new partnership formations and the shift towards greater

emphasis on ownership and *post-hoc* conditionality.[1] Benchmarks and indicators serve to monitor and govern by rendering the world legible, transforming, for example, different countries' complex socioeconomic and political features into numerical rankings.[2] Quite a lot is at stake in such benchmarking processes – both in producing and responding to them. As the editors note in their introduction to this Special Issue, benchmarks can neutralise or suppress political tensions and universalise particular interests through a quantitative and seemingly objective language.[3] The editors' conceptualisation of benchmarks invites us to reflect on the 'politics of benchmarking' and to assess how benchmarks work as tools for 'governing at a distance'.[4]

We follow the editors' focus on the politics of benchmarking and explore in more detail how benchmarks can be said to be or represent authority in global governance. We do this because extant literature on benchmarks has tended to attribute authority and power to global benchmarks without detailed analysis of what this implies for benchmarked actors. We specify what is meant by the concept of authority and foreground that it first and foremost describes a relationship between a superordinate and a subordinate actor, where the latter defers to the former. It follows that to assess the role and functioning of benchmarks, we have to look at the *uses* of benchmarks and at the reactions to them by benchmarked actors.[5] We draw on a case study of how the World Bank uses one of its benchmarks – the Country Policy and Institutional Assessment (CPIA) – in its 'governing' and relationship with two African countries, Ethiopia and Malawi.

Ethiopia and Malawi represent two interesting cases as both – but for different reasons – have experienced friction in their partnership relation with the World Bank, thus alerting us to local contestation and agency at the 'receiving end' over the global policy prescription embedded in the CPIA. The World Bank and its sister institution, the IMF, have disengaged their operations in both countries in certain periods, and both countries – though more so in Ethiopia – have had overt ideological disagreements with these institutions. The CPIA and the policy prescriptions that flow from it are more likely to be adopted in countries whose governments to a larger degree already share the World Bank's policy outlook. By focusing on two countries where the governments have resisted World Bank policy, and where the World Bank have discontinued budget support in certain periods, we can get a better grasp of the scope, and limits, of the authority of the CPIA as a benchmark, and of the World Bank as an organisation more generally.[6]

---

[1] For a good overview of this development, see Jacqueline Best, *Governing Failure: Provisional Expertise and the Transformation of Global Development Finance* (Cambridge: Cambridge University Press, 2014); Jon Harald Sande Lie, *Developmentality*: *An Ethnography of the World Bank–Uganda Partnership* (New York: Berghahn Books, 2015).

[2] Sally Engle Merry, 'Measuring the world: Indicators, human rights, and global governance', *Current Anthropology*, 52:3 (2011), pp. 83–95; James C. Scott, *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed* (New Haven: Yale University Press, 1998).

[3] André Broome and Joel Quirk, 'Governing the world at a distance: the practice of global benchmarking', *Review of International Studies*, 41:5 (2015), pp. 819–41.

[4] Nicholas Rose and Peter Miller, 'Political power beyond the state: Problematics of government', *British Journal of Sociology*, 43:2 (1992), pp. 173–205.

[5] Richard B. Friedman, 'On the concept of authority in political philosophy', in Joseph Raz (ed.), *Authority* (New York: New York University Press, 1990), pp. 56–91. For a discussion and application, see Ole Jacob Sending, *The Politics of Expertise: Competing for Authority in Global Governance* (Ann Arbor: University of Michigan Press, 2015), ch. 1.

[6] Lie, *Developmentality*; Graham Harrison, 'Post-conditionality politics and administrative reform: Reflections on the cases of Uganda and Tanzania', *Development and Change*, 32:3 (2001), pp. 657–79; Jon Harald Sande Lie, 'Challenging anthropology: Anthropological reflections on the ethnographic turn in International Relations', *Millennium*, 41:2 (2013), pp. 201–20.

We proceed as follows: We first present a brief discussion of the concept of authority, which prepares the ground for the empirical analysis of the CPIA. We focus on how the World Bank describes its role, including how it is used to allocate funds, to signal incentives to developing countries, and to work as a diagnostic tool for policy improvement. We then present findings from our two case studies of Ethiopia and Malawi. We focus on the role of the CPIA in allocating funds, in signalling incentives, and in shaping the policy dialogue and the Bank's country-specific strategy. We find that the CPIA is not – and contrary to what one would expect from the CPIA's nominal function and the literature on benchmarks – a very important tool for signalling incentives and allocating funds, or shaping the policy dialogue or the World Bank's strategy in these two countries. Rather, the CPIA is used highly selectively as one factor among many in the negotiations between World Bank staff and governments. These findings prompt us to reflect on the use of benchmarks and what type of authority that benchmarks such as the CPIA can be said to represent. We argue that the CPIA is best understood as a governing tool that the World Bank uses to govern and maintain its authority *vis-à-vis* both aid recipients *and* donors.[7] Presenting findings from our own study and that of others, we argue that the CPIA helps establish – together with other tools – the World Bank as an actor that embodies global authority on development, even if the CPIA (and other universalising governing practices) are less important as a tool in the World Bank's country-specific engagement, as in Ethiopia and Malawi, over the contents of government policies.

Authority and benchmarks

A key claim in extant literature is that benchmarks do, in fact, represent or contain power or authority. Davis *et al.* argue, for example, that: 'The burgeoning production and use of indicators in global governance has the potential to alter the forms, the exercise, and perhaps even the distributions of power in certain spheres of global governance.'[8] Alex Cooley notes that organisations that produce benchmarks and rankings can be 'judges' and 'regulators' and that benchmarks have the 'power to informally regulate global institutions and practices …'.[9] Similarly, Sally Engle Merry argues that indicators – what we here refer to as benchmarks – 'shift responsibility for governance from those in power to those who are governed'.[10] This focus is understandable, since most studies have analysed the 'ranking frenzy' as a novel macro-phenomenon where a range of actors uses benchmarks as a tool to govern and hierarchically order states.[11] Nonetheless, detailing the prevalence of benchmarking

---

[7] See Ole Jacob Sending and Iver B. Neumann, 'Banking on power: How some practices in an international organization anchor others', in Emanuel Adler and Vincent Pouliot (eds), *International Practices* (Cambridge: Cambridge University Press, 2011), pp. 231–54. See also Raymond D. Duvall and Arjun Chowdhury, 'Practices of theory', in Emanuel Adler and Vincent Pouliot (eds), *International Practices* (Cambridge: Cambridge University Press, 2011), pp. 335–54 (pp. 338–9).

[8] Kevin E. Davis, Benedict Kingsbury, and Sally Engle Merry, 'Introduction: Global governance by indicators', in Davis, Kingsbury, and Engle Merry (eds), *Global Governance by Indicators: Global Power through Quantification and Rankings* (Oxford: Oxford University Press, 2012), pp. 1–28 (p. 4).

[9] Alex Cooley, 'The emerging politics of international ranking and ratings', in Alex Cooley and Jack Snyder (eds), *Ranking the World: Grading States as a Tool of Global Governance* (Cambridge: Cambridge University Press, 2015), pp. 1–38 (p. 7).

[10] Sally Engle Merry, 'Measuring the world', pp. 83–95 (p. 90).

[11] Alex Cooley, 'The emerging politics of international ranking and ratings', pp. 9–13; Davis, Fisher, Kingsbury, and Engle Merry, *Global Governance by Indicators*.

as a form of governance at a global scale is not the same as demonstrating that benchmarks are authoritative in the sense of inducing deference and shaping the behaviour of benchmarked actors.

It should be recalled that authority designates first and foremost a specific type of relationship between actors. Crucially, most analysts agree that 'recognition' is a key component of authority. David Lake notes that authority 'does not exist without recognition',[12] and Deborah Avant, Martha Finnemore, and Susan Sell similarly note that 'Authority is created by the recognition, even if only tacit or informal, of others.'[13] This view of authority builds on Richard B. Friedman's discussion, where authority designates a relationship between a superordinate and a subordinate actor, which is *recognised*, and where the latter *defers* to the former.[14]

In this perspective, authority concerns the relationship between the producers/users of benchmarks, on the one hand, and the targets of benchmarks, on the other. Benchmarks are tools that can shape that relationship, but it depends on the context of its use. It also means that authority does not spring magically from some ideal-typical 'source' of authority, whether it is expertise, legal rules, or moral standing.[15] Rather, authority is something that is at stake in relationships where part of the competition revolves around what is to count as a 'mark' of authority.[16] While we think it makes sense to foreground the positions and resources of the actors involved in producing and using benchmarks (rather than analysing whether benchmarks themselves as having authority), we have to analyse how actors can use benchmarks to construe or strengthen a position of authority. This is not to say, however, that benchmarks can never be authoritative, only that to explain their authority, we must move beyond assertions about benchmarks' power to universalise and to substitute numerical descriptions for complex explanations, and explore the positions and power resources of those actors that go into establishing and using these benchmarks. In other words, what needs to be accounted for is not merely how a benchmark is made authoritative as a *description* or definition of some phenomenon, but also how that benchmark is *used* by different actors to try to shape the behaviour of their intended targets, and how intended targets of benchmarks *react* to these.

This has important implications for our empirical analysis, as we focus on how the actors involved use and respond to the CPIA in Ethiopia and Malawi. We also seek to be sensitive to the possibility that an actor that may recognise the epistemic authority of the World Bank, and of the CPIA as a benchmark, and yet decide to act in ways that go against such authority because of domestic political pressure or other considerations. Likewise, there is also the possibility that governments act in accordance with the benchmarks not because of the power of benchmarks but because the governments agree with the benchmarks' policy prescriptions.

But before we move on to analyse the uses of and reactions to the CPIA, we need first to present the CPIA and the rationale for its design and use as seen by the World Bank.

---

[12] David Lake, *Hierarchy in International Relations*, p. 8.

[13] Deborah Avant, Martha Finnemore, and Susan Sell (eds), *Who Governs the Globe?* (Cambridge: Cambridge University Press, 2010), p. 10.

[14] Richard B. Friedman, 'On the concept of authority in political philosophy', in Joseph Raz (ed.), *Authority* (New York: New York University Press, 1990), pp. 56–91 (pp. 60–2).

[15] Michael Zürn, Martin Binder, and Matthias Ecker-Ehrhardt, 'International authority and its politicization', *International Theory*, 4:1 (2012), pp. 69–106.

[16] See Ole Jacob Sending, *The Politics of Expertise*, ch. 1.

The formal order of the CPIA

The CPIA consists of 16 criteria grouped into four equally weighted clusters.[17] It is designed to measure a country's overall policies and institutional quality. It measures only those aspects that are under a country's control, and it focuses on actions and performance, not plans and intentions. By itself, the CPIA is merely one among many instruments the World Bank uses to assess partner countries. What makes the CPIA important is that it has always been tied to the allocation practices of the International Development Association's (IDA), which is the Bank's soft loan branch offering concessional lending on terms preferable to other types of loans offered by the Bank.[18] The CPIA provides the key input to the Bank's Performance Based Allocation (PBA) mechanism, where allocation of funds is based on 'needs' (measured by population size and poverty levels) and 'performance' (measured by the CPIA). The CPIA and the PBA together transforms a political question – how to allocate scarce resources from donors to recipients via the Bank – into an allegedly rules-based practice, based on the content and quality of countries' policies and institutions.

The CPIA matters for more than the allocations, however. It also signals incentives – partly via the PBA system – to client countries to adopt policies that conform to the policy prescriptions contained in the CPIA, to anchor the World Bank's 'Country Assistance Strategy', and to underwrite the World Bank's policy dialogue with governments of client countries.[19] The structural setup of the CPIA reflects the constitutive elements of benchmarking practices as identified by the editors of this Special Issue: *reification* by converting the diverse and subjective into legible, objective properties; *commensuration* by aligning the different units according to a common metric; *symbolic judgement* by ranking complex social, political and economic phenomena.[20] But it also goes beyond these in that it is directly linked to economic incentives, where higher CPIA scores translate into more funding through the PBA system.[21]

The CPIA consists of a set of criteria that represent the different policy and institutional dimensions that the World Bank deems important for effective poverty reduction and economic growth. The criteria have grown in number and scope, reflecting the institution's evolving priorities and policies. From an initial focus on financial management and economic indicators, it was expanded and made more comprehensive during the course of the 1990s. In 1997, governance was explicitly introduced through a new criterion on legal and regulatory framework. Here, concern with corruption was central, as evidenced by the reference to the importance of paying attention 'to the implications for corruption of public sector administration' in

---

[17] These clusters are: (A) Economic Management; (B) Structural Policies; (C) Policies for Social Inclusion and Equity; and (D) Public Sector Management.
[18] Established in 1960, IDA complements the World Bank's original lending arm, the International Bank for Reconstruction and Development (IBRD, established in 1944), thus signifying a focus shift from reconstruction to development and from middle-income to low-income developing countries. The IBRD and the IDA share the same management, staff and headquarters but finance their activities from different 'drawers'. IBRD lends at marked rates, functioning as a self-sustaining business providing loans to middle-income countries and credit-worthy poor countries. IDA lends money on concessional terms, that is, credits, meaning it charges little or no interest with a repayment period up to forty years, often including a five to ten years' grace period.
[19] See Elisa Van Waeyenberge, 'Selectivity at work: Country policy and institutional assessments at the World Bank', *European Journal of Development Research*, 21:5 (2009), pp. 792–810.
[20] Broome and Quirk, 'Governing the world at a distance'.
[21] Van Waeyenberge, 'Selectivity at work', p. 798.

assessing all elements – tax reforms, volume and composition of expenditures, public expenditure management, and civil administration'.[22] The following year, the CPIA was again expanded and refined as it now came to reflect 'an agreed set of Bank wide criteria for promoting growth and poverty reduction'.[23]

The World Bank started to disclose its ratings in the early 2000s. The reasons for disclosure are manifold, but for our purposes, two of them are particularly important. The first has to do with the World Bank's ambition to broaden the role and scope of the CPIA and improve its incentive structures. In order for the CPIA to creative incentives for compliance and have bearing on policymaking processes, the CPIA ratings have to be known. The World Bank has noted, for example, that it is important to simplify how the assessment in the CPIA factors into the PBA in order to 'make the system more effective at the country level'.[24] This is part of the reason for the Bank's decision to make the CPIA ratings public. As one review notes about the CPIA and PBA: 'Transparency and clarity are, of course, extremely important to the rationale for using a performance-based formula at all. The easier the formula is to understand, the more clearly can its incentive aspects be communicated to the eligible countries.'[25] The second has to do with the introduction of the Poverty Reduction Strategy (PRS) as a new mode of engaging with client countries, giving them more 'ownership' of policy formulation and implementation. This meant abandoning the Structural Adjustment Programmes (SAP) through which the World Bank imposed its readymade policies of privatisation and liberalisation onto multiple and diverse counterparts.[26] This approach, as part of the so-called Washington Consensus,[27] was highly criticised for being top-heavy and providing blueprint, one-size-fits-all models. The PRS-system, rolled out from 1999, set out to be more inclusive and context based, as recipient states now were to devise their own, national development strategy, based on principles of ownership and participation, to provide the foundation for the World Bank's financial support. In this context, the World Bank describes the role of the CPIA as a 'diagnostic tool' that 'indicate[s] areas where attention needs to be focused and thereby influence the focus of the IDA-supported program'.[28] Disclosing the CPIA, and connecting it to the incentive mechanism of performance-based allocations while widening its scope to signal a wider policy focus, can thus be seen as an instrument for the World Bank of combining greater freedoms and responsibilities in

[22] Operations Evaluation Department (OED), 'Review of the Performance Based Allocation System, IDA10–12' (Washington DC: The World Bank, 2001), p. 9.

[23] Ibid., p. 11.

[24] International Development Association, 'Report from the Executive Directors of the International Development Association To the Board of Governors' (Washington DC: World Bank, 2008), p. 21.

[25] Kenneth Watson, 'Mid-Term Review of the GEF Resource Allocation Framework – Comparison of the GEF RAF with other Performance-Based Allocation Systems', Technical Paper No. 8 (Washington DC: Global Environmental Facility, Evaluations Office, 2008), pp. 6–7.

[26] See, for example, Kate Bayliss, Ben Fine, and Elisa Van Waeyenberge (eds), *The Political Economy of Development: The World Bank, Neo-Liberalism and Development Research* (London: Pluto Press, 2011).

[27] 'Washington consensus', coined by Williamson in 1990, was originally a descriptive term denoting the least common denominator of policies pursued by Washington-based institutions such as the World Bank, the IMF, and the US Treasury Department to Latin America during the 1980s. The term comprised a set of ten specific policy prescriptions constituting the 'standard' reform package. Since then, the term has taken on a more normative meaning and is used in a different and broader sense, as synonymous with market fundamentalism, privatisation, liberalisation, neoliberal policies, deregulation of state institutions, and the intentionally expanding role of the free market. John Williamson, 'What Washington means by policy reform', in John Williamson (ed.), *Latin American Adjustment: How Much has Happened?* (Washington DC: Institute for International Economics, 1990).

[28] International Development Association (IDA), 'Summary of the Performance-Based Allocation System for IDA14' (Washington DC: World Bank, 2002), p. 6.

planning and implementation for client countries while at the same time securing an access point for governing by the World Bank.[29]

The production of the CPIA is done every year for each IDA eligible country by World Bank staff, that is, the Bank's respective country teams in cooperation with the region's Chief Economist. This is of some importance, as the actors engaged in assessing a country's performance are the same, at least at the country level, as those using the analysis when putting together the lending portfolio. The World Bank shares with the government the so-called 'write-ups' that underpins and justifies the numerical score given on each of the criteria contained in the CPIA. The government, however, is not involved in negotiating the benchmarking exercise and the World Bank's analysis.

The CPIA in practice: Incentives and policy dialogue in Ethiopia and Malawi

The World Bank presents the CPIA as performing two core functions: (1) allocate funds and signal incentives to partner countries for improved performance; and (2) shape the country-level policy dialogue, and the formulation of the World Bank's programmes in particular countries through the Country Assistance Strategy (CAS).[30] The fact that the World Bank identifies these functions as integral to the CPIA is an expression of the importance that the World Bank attaches to it in its management of relations with recipient governments. We focus on both allocations and incentives, and on the policy dialogue and the CAS because they touch on different aspects of the relations between the World Bank and the governments of Ethiopia and Malawi. The first centres on allocations and the signalling of incentives, where we would expect the CPIA to be a powerful tool to shape behaviour. The second centres on the CPIA's role in shaping the policy dialogue between the World Bank and in formulating the CAS, where we would expect the CPIA to play a different, and less direct, role in shaping the behaviour of the government than in signalling incentives and determining aid allocations.

Context: Ethiopia and Malawi

Ethiopia's relations with international donors date back to the 1950s. It was the first African country to receive World Bank assistance, in 1950. Today it tops the worldwide list of countries receiving aid from the US, UK, and the World Bank, receiving $3.5 billion on average in recent years, which represents 50 to 60 per cent of its national budget. As Ethiopia never was colonised, it had a state apparatus when it got engaged with donors. Thus it was quite unlike most other aid-recipient countries that only started to receive aid upon independence. One implication is that the Ethiopian authorities in past and present have seen aid relations as the meeting of equals, with a government more assured of its own policy directions, its entitlements

---

[29] Jon Harald Sande Lie, 'Developmentality: Indirect governance in the World Bank – Uganda Partnership', *Third World Quarterly*, 36:4 (2015), pp. 723–40. See also Best, 'Governing failure', ch. 5.

[30] Elisa Van Waeyenberge, 'Selectivity at work'. See also Julia Steets, 'Adaptation and Refinement of the World Bank's Country Policy and Institutional Assessment', Discussion paper commissioned by the Federal Ministry for Economic Cooperation and Development (Eschborn: GTZ and GPPI, 2008) and Independent Evaluation Group, 'The World Bank's Country Policy and Institutional Assessment: An Evaluation' (Washington DC: World Bank, 2009).

to the set development agenda and of its stature *vis-à-vis* the international donor community – as compared to most other aid recipient countries.[31]

Ethiopia has made development gains over the last decade in the health, education, and food security sector, and the national economy has seen a consecutive double-digit growth over the same period. Yet, Ethiopia still remains one of the poorest countries in the world: roughly 30 per cent live below the poverty line of $1.25 a day, and many are vulnerable to food insecurity. Over 80 per cent is dependent on subsistence agriculture and its fast growing population and climate change put greater pressure on natural resources and land, being the cornerstone of the country's growth. Ethiopia is seen to be a poor performer on governance issues, both according to the CPIA and other donor assessments. Yet, the same assessments deem Ethiopia to do well in terms of economic growth and service delivery. International donors agree, for example, that Ethiopia's good achievements are due to it having one of the most 'pro poor budgets' in Africa.[32] Ethiopia is therefore interesting inasmuch as it has established policies that go against the basic tenet of the CPIA, and indeed international development discourse in general, yet have succeeded in generating economic growth.

Malawi is one of the world's most densely populated and poorest countries. The population is at 13.6 million people, with around 80 per cent of the population living in rural areas.[33] By 2010 – the period where our empirical analysis in both countries was concluded – donors contributed around 40 per cent to Malawi's annual budget. In the period from 2004 to 2009, Malawi saw substantial economic growth, averaging 7.5 per cent. This economic growth period was preceded by a period of economic mismanagement where the IMF defined Malawi 'off-track', causing both the IMF and other donors, including the World Bank, to either suspend or drastically reduce their development assistance. Partly as a consequence of this reduction in development assistance, Malawi borrowed heavily in the private market, accumulating high levels of debt. In 2005, Malawi completed the IMF 'staff monitored program' and reached the completion point of the HIPC initiative in 2006, according it debt relief amounting to over US $2 billion.[34]

Corruption and financial mismanagement has been a key source of friction between the government and the World Bank (and other donors). One study argues that that mismanagement of public resources is so widespread that the negotiation over the government's budget is like a 'theatre': 'All the actors, from civil society, government, and donors seem aware that many of their statements and actions have little bearing on actual distribution of resources. Yet, all stakeholders 'act' as if the budget planning and formulation will actually have a bearing on the actual implementation and distribution of resources.'[35] As in Ethiopia, the government has clashed with the World Bank on major policy issues: in 2005, following years of drought and famine, President Mutharika introduced a programme for subsidies of

[31] Xavier Furtado and W. James Smith, 'Ethiopia: Aid, Ownership and Sovereignty', GEP Working Paper (Oxford: University of Oxford, 2007); The Oakland Institute, *Development Aid to Ethiopia: Overlooking Violence, Marginalization, and Political Repression* (Oakland: The Oakland Institute, 2013); Gérard Prunier and Éloi Ficquet (eds), *Understanding Contemporary Ethiopia: Monarchy, Revolution and the Legacy of Meles Zenawi* (London: Hurst, 2015).
[32] Interviews in Addis Ababa with officials and representatives of the Norwegian Embassy, 8 June 2009; UNDP, 9 June 2009; DFID, 10 June 2009; GTZ, 11 June 2009; and Swedish Embassy, 11 June 2009.
[33] Human Development Report (2009), available at: {http://hdr.undp.org/en/statistics} accessed 5 January 2010.
[34] {http://lilongwe.usembassy.gov/economic_commercial_section2.html}.
[35] Lise Rakner et al., 'Budget as theater', p. iv.

fertilisers to boost agricultural production against the recommendation of the World Bank and other donors.

Allocations and incentives

From 2005 to 2008, Ethiopia's annual, average CPIA score was 3.4, on a scale from 1 to 6, which is quite low (states with an average below 3.2 are categorised as fragile).[36] In 2005, moreover, the World Bank and other donors were highly critical of the government's handling of the post-2005 election turmoil, whereupon several donors reduced their support. This political situation caused the World Bank to terminate its budget support mechanism, which is difficult to interpret as anything else than the World Bank signalling discontent with the government's handling of the situation. Nonetheless, we see no cuts in the World Bank's overall allocations to Ethiopia in this period. Rather, IDA funding increased from US $449.9 million to 634.5 million over the same period.[37] This increase in allocation occurred at the same time as the World Bank moved from direct budget support into more direct programme support. Since IDA allocations are determined annually, with the CPIA score as the main determinant, this is a strong indication that the CPIA does not determine aid allocations in a way that clearly signals incentives to behave in keeping with the prescriptions of the CPIA. Part of the explanation for this decoupling between performance assessment on the one hand, and actual allocations, on the other, is that other factors kick in and shape the allocation to each country, including changes in 'performance relative to other countries, changes in eligibility for IDA resources and for IDA grants, and availability of IDA resources'.[38] We mention this in some detail because it raises the issue, discussed in more detail below, that we should be attentive to how the World Bank presents itself and its operations with a view to securing its credibility and authority *vis-à-vis* both recipients and donors at the same time.

From 2005 to 2008, there was a significant decrease in the CPIA score on macroeconomic management, from 3.5 in 2005 to 2.5 in 2008. According to the World Bank, Ethiopia's growth is challenged by high domestic inflation and a difficult balance of payment situation with weak foreign exchange reserves driving fuel and food prices up. In order to mitigate and control the high domestic inflation rates, the government has removed fuel price subsidies, and the National Bank has raised the minimum reserve requirement on commercial bank deposits. Nevertheless, the World Bank asserts that the macroeconomic situation remains tenuous,[39] and the decrease in CPIA score on economic management is made with reference to an incomplete sequencing and implementation of what the Bank sees as a sound policy response by the government.[40] And yet, the Ethiopian economy experienced strong and broad based growth over the last decade, averaging 10.9 per year in 2003 to 2012 compared

---

[36] While the rounded average for 2005 and 2006 is 3.4, the actual average for 2007 is 3.425 and 3.35 for 2008.

[37] These figures derive from the respective and relevant CAS. Available at: {http://go.worldbank.org/ZBCG9IF761}.

[38] See 'How IDA Resources are Allocated', available at: {http://www.worldbank.org/ida/how-ida-resources-allocated.htmlAnnex 2} accessed 17 September. Quotation is from 'Annex 2 IDA's Performance Based Allocation System for IDA17', p. 2.

[39] Interview with senior official at World Bank country office in Addis Ababa, 11 June 2009.

[40] The Bank recognises that these macroeconomic challenges are exacerbated by a steep rise in global food and commodity prices, the international financial crisis, and the lack of rain during the first half of 2008. While influencing the domestic economy, these are all exogenous factors outside the control of government and thus they are nominally not part of what the CPIA assessment is expected to take into account.

to the regional average of about 5 per cent, producing positive trends in reducing poverty in urban and rural areas. Among critics of the World Bank's use of the CPIA to determine aid allocations, this is a key point: Many countries, including Ethiopia, have been able to generate stable economic growth and poverty reduction without conforming to the prescriptions contained in the CPIA.[41] All interviewees agreed that these measures were undertaken on the government's initiative to improve the macroeconomic situation, and not – by their own admission – as an attempt to improve the CPIA rating. Both government and donor representatives also report that the government would never alter its policies just for the sake of accommodating donors or to increase its CPIA to gain more IDA funding. Instead, interviewees held that the government would rather opt to lose development assistance than accommodating donors' demands and policy prescriptions.[42]

A similar pattern can be found in Malawi, where CPIA ratings have been quite stable over time, with some significant changes on the cluster on Economic Management, where fiscal policy improved from 3 to 3.5 between 2006 and 2007, and macroeconomic management also improved from 3 to 3.5 from 2005 to 2006. IDA allocations, however, have changed quite significantly in a negative direction in the same period, from US $116 million in 2005 to 47.2 million for 2006. The CPIA ratings cannot explain such large changes from one year to another. *Ceteris paribus*, Malawi would receive a higher allocation with an improvement in CPIA rating. As with Ethiopia, both primary and secondary sources suggests that even in the case of a strong link between performance and allocations, it is unlikely that it would contribute to Malawi changing its policies or institutions in a certain direction.[43] A case in point is the decision to introduce subsidies of fertilisers against the explicit recommendation of the World Bank. As in Ethiopia, the government appeared to have made headway in economic terms by going against the policy prescriptions of the World Bank. As reported by the New York Times:

Over the past 20 years, the World Bank and some rich nations Malawi depend on for aid have periodically pressed this small, landlocked country to adhere to free market policies and cut back or eliminate fertilizer subsidies, even as the United States and Europe extensively subsidized their own farmers. But after the 2005 harvest, the worst in a decade, Bingu wa Mutharika, Malawi's newlyelected president, decided to follow what the West practiced, not what it preached.[44]

Interestingly, the government's decision to subsidise fertilisers did yield tangible results in corn production, making Malawi a net exporter of corn rather than having to rely on food aid from donors in a matter of years. In response to this decision, the

---

[41] Ravi Kanbur, 'Reforming the Formula: A Modest Proposal for Introducing Development Outcomes in IDA Allocation Procedures', D:scussion Paper No. 4071, Centre for Economic Policy Research, London (2005).

[42] This view is reflected by just about all interviewees, for example, a senior official at MoFED's Multilateral Cooperation Department, 12 June 2009; World Bank senior officials, 10 June 2009; a political affairs officer at the British Embassy, 11 June 2009; UNDP representative, 10 June 2009. See also Axel Borchgrevink, 'Limits to donor influence: Ethiopia, aid and conditionality', *Forum for Development Studies*, 35:2 (2008), pp. 195–220.

[43] See Goran Hyden, *African Politics in Comparative Perspective* (Cambridge: Cambridge University Press, 2006); Stein Sundstøl Eriksen and Kristen Nordhaug, 'Politics of survival in the making of strong and weak states', *Forum for Development Studies*, 33:2 (2005); Borchgrevink, 'Limits to donor influence'.

[44] Celia W. Dugger, 'Ending Famine, Simply by Ignoring the Experts', *New York Times* (2 December 2007), available at: {http://www.nytimes.com/2007/12/02/world/africa/02malawi.html?pagewanted=print&_r=0} accessed 10 August 2014.

World Bank reduced its support from US $40 to 30 million to signal discontent with the policy, in coordination with other major donors. This did not entail a large overall reduction in aid, however, as it was reportedly a coordinated move: 'When the Bank is not happy about something, it signals this by reducing its support somewhat, but it allows the other donors to continue its support. It's a bad cop, good cop thing.'[45]

The lack of a strong or direct link between the CPIA and aid allocations, and also between the policy prescriptions contained in the CPIA and government policy, suggest that the CPIA is not authoritative. If the CPIA had been authoritative, we would expect the governments in Ethiopia and Malawi to 'defer' to its policy prescriptions, and the economic incentives associated with the CPIA – via the PBA – would have weighed heavily on the governments' decision-making. While we should be cautious about over stating our case and note that the CPIA may well induce deference in other countries, our findings are not unique: Other studies have documented that IDA allocations may change significantly from year to year, due to political pressure from key donors in the Bank. Andersen *et al.* note, for example, that the World Bank decided to withdraw lending to Allende's Chile in 1970–3; and during the Iran-Iraqi war, the IDA stopped funding to both countries. Moreover, World Bank assistance to Pakistan tripled from US $226 million in 2001 to $860 million in 2002.[46] Let us now turn to the less direct yet potentially equally important role of the CPIA as a tool that frames the policy dialogue and informs the World Bank's country-specific operations via the Country Assistance Strategy.

Policy dialogue and country strategy

The CPIA aspires to structure the contents of the policy dialogue at the country level. In Ethiopia in April 2009, at a high-level meeting between government ministers and donor representatives from the World Bank and the IMF as well as bilateral donors,[47] the macroeconomic situation and the newly-ratified civil society organisations (CSO) law were up for discussion. While creating concerns among donors and squarely being part of the CPIA's governance benchmarks, the World Bank and the IMF did not comment on the political sides of the CSO law in this High Level Forum. Rather, they commented on the macroeconomic issues where the IMF representative focused on topics falling under CPIA's Economic Management cluster, and the World Bank representative attended to issues dealing with structural policies, integral to cluster B of the CPIA.[48]

The World Bank representative addressed the structural policies, being central to the CPIA benchmarks, in making a call for stabilising the economy by modernising the financial sector to facilitate private sector activities. This would involve opening up and privatising key institutions such as the banking system and the telecommunication sector, and open up for private ownership of land. This is a case of World Bank staff using the CPIA as a benchmark, where the contents of one government's policy is held up against that of other countries and the 'best practice'

[45] Interview, Lilongwe, 26 August 2009.
[46] Thomas B. Andersen, Henrik Hansen, and Thomas Markussen, 'US Politics and World Bank IDA Lending,' Discussion Papers 05–06 (University of Copenhagen, 2006), p. 3.
[47] These donors comprise the Development Assistance Group (DAG) in Ethiopia, which is a mechanism for coordination among bi- and multilateral donors.
[48] Interview with UNDP official, 9 June 2009. See also minutes from Ninth Government–Donor High Level Forum, Monday 20 April 2009. Ministry of Finance and Economic Development.

implied by the CPIA. At the meeting, however, the government clearly signalled discontent with the World Bank's policy prescription, and clearly stated that it did not agree: it did not want to liberalise the financial sector, open the banking system to foreign investment, privatise telecommunication, or privatise land ownership. Indeed, the Ethiopian government challenged the contents of the CPIA for reasons of ideology, its alleged subjectivity, and its lack of transparency – seeing it as an ideological package promulgated by the World Bank.[49] The government sees the debate over structural policies as a form of imposition directly challenging its own ideological orientation.[50] For example, the government's celebrated 'pro-poor budget' is, amongst others, based on revenues derived in part from its ownership of telecommunication, land, and the banking system – sectors the World Bank would like to see privatised. Consequently, the CPIA is criticised for rewarding policies seen neither appropriate nor desirable by the Ethiopian government. As in Malawi (see below), the government was willing to discuss macroeconomic management, which it itself deems relevant as it directly impacts in creditworthiness, but not governance issues relating to privatisation, which it deems to be ideological and political.[51]

The contestation over ideology relates to the critique over CPIA's focus on performance, and indicates that the CPIA is not authoritative and hence does not induce deference from the Ethiopian government. The World Bank, for its part, seems to go out of its way to position itself so as to have some say on and role in shaping government policy. For example, the World Bank has called for the resumption of regular budget support even though the political reasons for terminating it is still in place, because the World Bank considers the annual policy negotiations they allow for as far more influential than what is provided for by the CPIA. This suggests that the World Bank is seeking to leverage its position to shape governmental policy, using the CPIA as a tool to manoeuvre and establish a position from where to engage the government on how it should govern. As such, the CPIA is one tool among many that the World Bank uses in its continual negotiation with and efforts to shape the behaviour of, the Ethiopian government.

In Malawi, the case is somewhat different, since there is not the same type of explicit anti-Bank sentiment or ideology within the government. Interestingly, however, the CPIA is known to only a handful of people inside the Ministry of Finance. The responsible section within the Finance Ministry (Debt and Aid) in Malawi has tried to broaden the understanding and implications of the CPIA and the PBA across ministries, reportedly telling colleagues in other ministries that: 'There is no room for complacency' as the CPIA ratings affect aid allocations from IDA.[52] Nevertheless, only those that deal directly with Bank officials and that coordinate with donors are aware of what the CPIA is and why it matters. Because Malawi received budget support in this period, in contrast to Ethiopia, the central arena for policy dialogue was the consultations between the government and the donor group, which includes the World Bank and bilateral donors. This donor group has established a Performance Assessment Framework with detailed targets, actions, and indicators of progress. While the indicators and targets broadly reflect the type of

---

[49] Interviews, MoFED, 9 and 12 June 20009 and with a DFID programme officer, 10 June 2009.
[50] Interviews, MoFED, 12 June 2009. This sentiment is also acknowledged by Bank representatives (interviews at the Bank's country office), 10 June 2009.
[51] Interview, MoFED-officials, 9 June 2009.
[52] Interview, Officials with the Ministry of Finance, Lilongwe, 26 August 2009.

development thinking that underwrites the CPIA, the CPIA as such does not play any significant role in ongoing debates between donors and the government.

The CPIA therefore seems to be central to the discussions between Bank officials and a small group of officials inside the Ministry of Finance, but it is not central to the policy dialogue in broader terms. Bank officials also acknowledged this, noting that: 'Ideally, government should use the CPIA but they don't.'[53] This view was also communicated by officials in the Ministry of Finance, with one informant noting that: 'We [Debt and Aid in the Ministry of Finance] have tried to sell CPIA to other parts of the government, saying that it matters, but with limited effect.'[54] The lack of 'traction' and even information about the CPIA in Malawi are also reflected in the World Bank's country specific strategy. World Bank support to Malawi was distributed with 29 per cent for human development, 34 per cent for infrastructure, 24 per cent for agriculture, rural development and natural resources, and so on.[55] This prioritisation dovetails with the *government's* priorities, but it does not harmonise with the CPIA ratings, indicating that the Bank aligns itself with the government rather than with the contents of its own assessment. Or rather, that the World Bank balances the policy implications from the CPIA assessment against the government's priorities.

Thus far, we have focused on the role of benchmarks in two particular settings, and have assessed whether the CPIA can be said to help the World Bank induce deference. While we see indications that the World Bank is authoritative as regards macroeconomic issues, it is not on other aspects, including liberalisation of key sectors of the economy (Ethiopia), and the use of subsidies (Malawi). Moreover, the CPIA does not appear to be a central tool in the World Bank's policy dialogue with the government, nor in the formulation of the World Bank's country-specific priorities. The picture that emerges is one where the World Bank uses the CPIA as one tool among many in its continual negotiation with, and efforts to shape the policies of, these governments. What to make of this? Should we conclude that the CPIA, as a benchmark, have no noticeable effect in these settings, and that the World Bank is not authoritative? Not necessarily. Other governments may relate to the CPIA in a way that is very different from Ethiopia and Malawi. As we argue below, however, the key attributes of benchmarks make them useful for actors that claim a particular type of global or universal authority, but by virtue of this claim to universal application, based on a particular view of how societies should be governed, these benchmarks are simultaneously vulnerable to contestation by actors that have concrete authority in particular settings. Again, we turn our attention to benchmarks as a governing tool, but we now broaden our focus and include in our discussion what the CPIA do to the relationship between the World Bank and donors.

What benchmarks do: Universal vs concrete authority

Benchmarks tend, as the editors of this Special Issue note, to cloak the prescriptive elements that they contain via numerical expressions and can present as description what is prescription. More specifically, benchmarks may constitute what Michel

---

[53] Interview, World Bank Official, Lilongwe, 24 August 2009.
[54] Interview, Officials with the Ministry of Finance, Lilongwe, 26 August 2009.
[55] Malawi Country Brief, World Bank. Updated June 2009, available at: {http://web.worldbank.org/WBSITE/ EXTERNAL/COUNTRIES/AFRICAEXT/MALAWIEXTN/0,,menuPK:355882~pagePK:141132~piPK: 141107~theSitePK:355870,00.html} accessed 13 September 2009.

Callon and Bruno Latour call an 'inscription' – the translation of particular interests into 'material form', or more generally making that which is contested into something uncontested and immutable that can circulate and gain the appearance of objective and fixed form.[56] Benchmarks help in making some ideas more or less uncontested and immutable, which circulate and gain the appearance of objectivity. Through processes of reification, commensuration, and symbolic judgment, benchmarks give an appearance of facticity to ideas and enable their use by a broad range of actors as they engage in global governance.[57]

The CPIA is a case in point. The core causal claim behind the substantive contents of the CPIA is that a good policy environment is a necessary condition for the effective use of aid. This idea came to prominence with the highly influential study by two World Bank researchers on 'Aid, Policies, and Growth'.[58] The CPIA helps produce an image of this causal relationship as objective and uncontested not only by making it numerical and creating a comparable assessment of the quality of each country's policy environment. It also renders possible a broad range of other activities, such as making allocations according to a rules-based system (the PBA), and creating an infrastructure of seemingly robust data about whether developing countries' policies are improving that donors and others use in debating the contents of World Bank policy.[59] The continual *practical* use of the CPIA by the World Bank and others adds facticity to the data that it contains, even if the causal claim that it builds on is contested in the scholarly literature. With Alain Desrosieres we can say that the CPIA 'furnish[es] something to lean on, or at least a common language of reference, around which the famous "democratic debate" can be organized'.[60]

The CPIA fills such a function within the World Bank as the indicators and rankings it produces – even when they are disputed and negotiated – serve as the common framework where it is possible to agree and disagree about how to guide, assist, and help developing countries.[61] For example, there is considerable debate within IDA about the weight of 'performance' relative to 'needs' in allocating funds. This debate is driven by distinct political views of what constitutes 'development' and how it should be best promoted. The PBA system, made possible by the CPIA, here defines the framework within which every actor has to formulate their position. Rather than debating, say, whether country x is to receive more funds than country y, the debate operates at the level of the relative weight of needs to performance in the PBA formula, with the CPIA being a largely unproblematised input to this formula.

Note that the CPIA is here important first and foremost in how the World Bank manages its relations with donors. The World Bank is dependent on funds from donor governments, and the forerunner to the CPIA, established in the late 1970s, focused mainly on recipient governments' capacity to manage funds: It was established as a governing tool to address donors' fiduciary concerns over how the funds they contributed to IDA was ultimately managed and spent at the country level.

[56] Michel Callon, 'Techno-economic networks and irreversibility', in John Law (ed.), *A Sociology of Monsters: Essays on Power, Technology and Domination* (London: Routledge, 1991), pp. 132–61 (p. 143); Bruno Latour, *Science in Action: How to Follow Scienstists and Engineers through Society* (Cambridge: Harvard University Press, 1987).

[57] Broome and Quirk, 'Governing the world at a distance'. See also, Pierre Bourdieu, 'The mystery of the ministry: From particular wills to the general will', *Constellations*, 11:1 (2004), pp. 37–43.

[58] Craig Burnside and David Dollar, 'Aid, Policies and Growth', Working Paper No. 1777 (Washington DC, World Bank, 1997).

[59] Sending and Neumann, 'Banking on power'.

[60] Alain Desrosieres, 'For a politics of the tools of knowledge', p. 1.

[61] Sending and Neumann, 'Banking on power'.

As such, the CPIA is as much about 'governing' relations with donors as with recipients as it addresses donors' concern with performance and fears of misuse of funds and corruption. Moreover, the CPIA enables the World Bank to bring the whole gamut of recipient countries' domestic policies into its purview without seemingly coming into conflict with its ostensibly apolitical mandate.[62] It does so by transforming the complex sociopolitical and economic structures of recipient countries into numbers on a total of sixteen indicators, thereby presenting as description what is prescription, and as eminently governable by the tools at the hands of the Bank. The CPIA is in this sense used by the World Bank to (re)produce an image of the Bank as an actor capable of governing and that donors can entrust with their funds in an effort to guide recipient governments towards 'development'.

This role of the CPIA is distinct from the role it plays in particular borrowing countries. As we saw above, the CPIA was used by the World Bank to try to induce deference in the governments of Ethiopia and Malawi to follow its policy prescriptions on specific policies. The role of the CPIA in managing relations with donor governments is of a different type, having to do with the establishment of a credible image of the World Bank as an international organisation that run on rules and expertise, rather than on the power and political preferences on some key states. Here, the CPIA helps in establishing a different type of authority, one that operates in a universal register, where some rules and norms are held to be good or effective or necessary everywhere, regardless of the particular context.[63] This is of crucial importance: International organisations and non-governmental organisations (both for-profit and non-profit) establish benchmarks that are typically global in scope, targeting governments, firms, and other actors. Thus understood, benchmarks can perhaps best be seen as institutionalising a universalist language, where some categories and indicators are effective in shaping global policy discourse and presenting as public interests that which is particular interests.[64] Benchmarks here construct a particular vision of the world and of the objects to be governed, enabling 'governing at a distance'.[65]

But precisely because the CPIA operates in such a universalist register, used by the World Bank to establish what amounts to an Archimedean point above any specific location or particular context,[66] it is not tailored to grasp what may produce economic growth in any particular setting. Granted, some benchmarks may be authoritative in particular settings because they are universal. Human rights norms may, for example, be used in particular settings to effect policy change with reference to a universal norm. Their authority to effectuate policy change would nonetheless depend on the configuration of political forces in the context where they are to be applied.[67] The CPIA, and other benchmarks that implicitly make causal claims about what works, would have to attain the quality of a universal and law-like mechanism in order to overcome the possible tension between the universal and the particular: the

---

[62] Julia Steets, 'Adaptation and Refinement of the World Bank's Country Policy and Institutional Assessment', Discussion paper commissioned by the Federal Ministry for Economic Cooperation and Development (Eschborn: GTZ and GPPI, 2008).

[63] Stephen Toulmin, *Cosmopolis: The Hidden Agenda of Modernity* (Chicago: University of Chicago Press, 1992).

[64] Stein Sundstøl Eriksen and Ole Jacob Sending, 'There is no global public: the idea of the public and the legitimation of governance', *International Theory*, 5:2 (2013), pp. 213–37.

[65] Rose and Miller, 'Political power beyond the state'.

[66] On this, see Nehal Bhuta, 'Against the state-building', *Constellations*, 15:4 (2008), pp. 517–42.

[67] See Thomas Risse, Stephen C. Ropp, and Kathryn Sikkink (eds), *The Power of Human Rights: International Norms and Domestic Change* (Cambridge: Cambridge University Press, 1999).

laws of physics do apply everywhere, thereby making every particular instantiation a confirmation of the universal. But the CPIA builds on one among many different and competing theories or claims about the drivers of development, which makes its claim to universality vulnerable in the face of evidence that does not fit. As former World Bank researcher Ravi Kanbur has noted, the CPIA measures policies, not development outcomes, and so it cannot account for countries that have poor policies according to the CPIA standard, yet manages to produce high growth rates.[68] We saw this above in the analysis of Ethiopia and Malawi, both of which have had a level of success in generating growth and poverty reduction (Ethiopia) or boosting agricultural production to become a net exporter (Malawi) through policies deemed inappropriate in the CPIA. This helps explain why the CPIA is not necessarily a significant input factor when a national government makes policy decisions, as such decisions are the outcome of national political negotiations to which external actors are often not privy. As Ann Swidler has noted, 'Despite donors' prestige and financial heft, they have more difficulty penetrating and altering local patterns of governance than one might expect. … [W]hat donor organizations offer is received (or seized) within a different social organization, where intended and actual effects differ.'[69] What makes the CPIA effective in establishing the World Bank with a level of global authority on development, then, hails in no small part from reading out variation between countries by setting up a standard for behaviour that invariably reflect a particular view of society.[70] When such a standard is used to discuss a general vision for governance and to set terms for resource allocation, it does not interfere with any particular country's substantive policies on, say, privatisation, social policy, or taxation. But when it is applied in particular contexts and used as a tool to shape governments' policy decisions, it potentially does come into conflict with domestic political preferences, thereby undermining its efficacy as tool to shape particular countries' policies.

One way to understand this is in terms of Stephen Hopgood's discussion of the tension between claiming authority with reference to non-contingent principles applicable everywhere, advanced and defined by cosmopolitans that work (for example) for international organisations, and the 'claims to national and ethnic loyalty, to historical cultural traditions'. The former is thin, associated with the 'stranger', the latter is thick and linked to belonging and substantive identities.[71] There is therefore quite a lot at stake in how benchmarks are produced and used by organisations such as the World Bank. The CPIA arguably augments the generalised or universal authority of the World Bank, which is important in its relationship with donors. By the same token, however, the CPIA – and benchmarks more generally – help establish a type of relationship with the object to govern that is 'thin' and therefore less able to understand and act on the nitty gritty elements in particular contexts.

We should not downplay what this amounts to, since it means that as the CPIA and other universalising tools become the key register for debates about development, and the concrete settings are not considered sources of meaning and policy guidance

---

[68] Ravi Kanbur, 'Reforming the formula: a modest proposal for introducing development outcomes in IDA allocation procedures', *Revue d'Economie du Developpement* (2005), pp. 79–99.

[69] Ann Swidler, 'Dialectics of patronage', in David Hammack and Steven Heydemann (eds), *Globalization, Philanthropy, and Civil Society: Projecting Institutional Logics Abroad* (Bloomington: Indiana University Press, 2009), p. 197.

[70] Andre Broome and Leonard Seabrooke, 'Seeing like an international organization', *New Political Economy*, 17:1 (2012), pp. 1–16.

[71] Stephen Hopgood, 'Moral authority, modernity and the politics of the sacred', *European Journal of International Relations*, 15:2 (2009), pp. 229–55 (p. 238).

in their own right. Rather, just as with indexes and benchmarks of fragile states, they impose categories that capture what is lacking rather than what is already there.[72] International development professionals may represent different actors, be they states, international organisations, non-governmental organisations, business or more loosely organised groups. They nonetheless share a universalised language of indicators and benchmarks, and have professional stakes in being recognised as competent participants in debates about development.[73] But this is quite distinct from the task of transforming, 'local patterns of governance'.[74] We are now in a position to better understand the role of the CPIA. In Ethiopia and Malawi, the CPIA is officially seen by the World Bank as a central policy tool. And yet, in both cases, other factors – whether it is long-established ideological grounds (Ethiopia) or disagreement on specific policy choices (Malawi) – take precedence over the policy prescriptions that are grounded in claims to universality: In both countries, political leaders directly challenged (Ethiopia) or sought to bypass (Malawi) the policy prescriptions of the CPIA. In Ethiopia, reduced CPIA scores and discontinuation of budget support occurred at the same time as an increase in IDA allocations. This apparent paradox can be understood if we consider that the World Bank is using the CPIA to manage relations with donors and to stake a claim to universal authority on development issues. At the same time as the World Bank is seeking to secure this position of a rules-based organisation, which delivers the same policy advice to all aid recipients, however, it also has to stay engaged and seek to have some level of influence in particular settings, which calls for deviation from one-size-fits-all prescriptions. By reducing CPIA scores and moving away from budget support while at the same time increasing IDA allocations, the World Bank manages its relations with donors *and* recipients. The same dynamics is on display in Malawi, where the World Bank coordinated with other donors to send a signal to the government that they were critical of government policy, thereby cutting support, yet maintained working relations to seek influence on other policy areas.[75] Indeed, staying engaged and maintaining the partnership relation seems to trump the role of incentives and performance that the World Bank use to describe its operations publicly. The fact that a donor needs a recipient just as a recipient needs a donor illustrates what is often downplayed – that the World Bank, as any bank, is dependent on upholding its lending portfolio and market shares *vis-à-vis* other donors. The donor and client parties' joint reluctance to frustrate the overall partnership gives both parts some leverage on the sly of benchmarks' nominal order.[76]

Conclusion: the limits of global authority

Benchmarks engender a particular type of authority, which is universal in scope, but for this reason it is not necessarily effective in concrete settings. As such, benchmarks

[72] Stein Sundstøl Eriksen, '"State failure" in theory and practice: the idea of the state and the contradictions of state formation', *Review of International Studies*, 37:1 (2011), pp. 229–47.

[73] Didier Bigo, 'Pierre Bourdieu and International Relations: Power of practices, practices of power', *International Political Sociology*, 5:3 (2011), pp. 225–58.

[74] Ann Swidler, 'Dialectics of patronage', in David Hammack and Steven Heydemann (eds), *Globalization, Philanthropy, and Civil Society: Projecting Institutional Logics Abroad* (Bloomington: Indiana University Press, 2009), p. 197.

[75] Interview by telephone with World Bank official, 18 August 2009.

[76] Lie, *Developmentality*.

are examples of the limits of global authority, and of a mismatch between the register used to describe and explain global challenges such as poverty, on the one hand, and the concrete strategies that may be effective in shaping behaviour in the particular contexts, on the other. We should not, however, conclude from this that benchmarks are not important in particular contexts such as the one under analysis here. As we saw above, the CPIA helps establish the parameters within which policies are formulated and justified as recipient governments engage the World Bank and other donors: In both Ethiopia and Malawi, the World Bank criticised the government's managing of its economy, and the basis for that critique was taken from the standard defined by the CPIA. Our point is that the CPIA – because it operates in a universal register – is important in establishing the basis of judgment and criticism, but because of this, it is not a very relevant tool for the concrete efforts to shape behaviour or guide policy, where a range of contextual factors (political and economic) tend to dominate.

Against this backdrop, we may ask what benchmarks tell us about the actors that produce them, and whether benchmarks are not part and parcel of an 'audit culture' where a particular language or discourse of performance, effectiveness, and accountability is established.[77] Benchmarks fit what John Meyer and his colleagues call 'rationalized myths' – scripts that are enacted to confirm actors' status as 'modern', 'rational', 'accountable', and so on.[78] Such a conceptualisation of benchmarks is pertinent, we believe, because it helps explain their prevalence as a technology of governing at the level of global policy discourse even when it appears of little use in concrete settings to address the problems it is purportedly designed to address. Performing or engaging in a particular practice – such as benchmarking – helps produce the image of an actor as competent and authority qua participant in global governance.[79] A similar view has been advanced among students of development practice. For David Mosse, for example, policy (or benchmarks) should not be seen as coming before and shaping practice – as if an actor's identity is a given and on the basis of which policy decisions are made and practices subsequently emerge.[80] Rather, Mosse invites us to see how 'actors in development devote their energies to maintaining *coherent representations* regardless of events'.[81] The CPIA is hailed as important, and defines the infrastructure for the World Bank's management of relations with donors. It is also used to allocate funds, even if the 'rule' of performance based allocations is not always followed. At the country level, the CPIA is said to signal incentives and to inform the policy dialogue, but in practice it is one policy tool among many that the World Bank deploys in its efforts to position itself to have a dialogue with the government to try to shape policy.

---

[77] Marilyn Strathern (ed.), *Audit Cultures: Anthropological Studies of Accountability, Ethics and the Academy* (London: Routledge, 2000).

[78] John W. Meyer and Brian Rowan, 'Institutionalized organizations: Formal structure as myth and ceremony', *American Journal of Sociology*, 83:2 (1977), pp. 340–63.

[79] Raymond D. Duvall and Arjun Chowdhury, 'Practices of theory', in Emanuel Adler and Vincent Pouliot (eds), *International Practices* (Cambridge: Cambridge University Press, 2011) pp. 335–54 (p. 339).

[80] David Mosse, *Cultivating Development. An Ethnography of Aid Policy and Practice* (London: Pluto Press, 2005); Lie, *Developmentality*.

[81] Mosse, *Cultivating Development*, p. 2, emphasis added.

ARTICLES

British International
Studies Association
BISA

MIX
Paper from
responsible sources
FSC™ C013985
FSC
www.fsc.org

CAMBRIDGE
UNIVERSITY PRESS