

Learning from Experiments in Optimization: Post-Critical Perspectives on Monitoring and Evaluation¹

Brit Ross Winthereik, Associate Professor, IT University of Copenhagen
Casper Bruun Jensen, Associate Professor, Osaka University

Abstract

This article examines attempts by professionals in the Danish branch of the environmental NGO NatureAid to optimize their practice by developing a local standard. Describing these efforts as an experiment in optimization, we outline a post-critical alternative to critiques that centre on the reductive effects of management and audit. The notion that reduction is inherently negative fails to recognize that achieving specific forms of reduction is often the reflexive aim of standardization. Rather than resisting monitoring and evaluation, the environmental consultants we study try to create a system capable of constraining their work in the *right way*.

Focusing on this experiment in optimization allows us to redescribe audit as a varied set of practices and aspirations, embedded in standards that generate relative, forms of organizational transparency and opacity. This offers a view of management as ‘broken up;’ as a distributed, ambient activity, variably performed by different actors using different standards.

Keywords:

Audit, Monitoring and evaluation, Experimentation, NGOs, Optimization, Performativity, Post-critical ethnography, Standardization, Transparency

Acknowledgments: We are grateful to the *NatureAid* professionals who let us study their work. We would like to thank the guest editors for encouraging us to submit the paper, and for their suggestions for improvement. We appreciate the thoughtful comments and recommendations by the anonymous reviewers.

Introduction

Western societies increasingly document just about any activity – from children’s reading abilities, to patterns of consumption and the accomplishments of organizations – and render them subject to evaluation. According to the sociologist and management scholar Michael Power, by the mid-1990s evaluations were so widespread that one could speak of an audit explosion (Power 1994, 1997; Jensen and

¹ This paper is a modified version of Winthereik and Jensen (2014), originally published in Danish.

Winthereik 2013, 121-147). Power further argued that evaluations gradually gained a legitimizing, or even ritual, function in modern societies. To a significant extent, they came to be conducted for their own sake. The consequence was that the assumed link between evaluation, quality improvement and optimization was often tenuous, if not altogether missing. Often, Power argued, evaluation resulted in little but new forms of 'learned ignorance' (Power 1997, 123).

In this paper, we focus on optimization efforts in the Danish office of the international NGO *NatureAid*.² Because *NatureAid* is concerned with changing socio-environmental conditions in developing countries, this is a place where the concerns of critical management studies with the negative consequences of standardization and optimization intersects with the critique of international development formulated by the anthropology of development. Focusing on standardization as an experimental practice, we offer a description of audit not as a system of domination but as a set of varied practices, the interaction of which create relative transparencies and opacities. Generating an image of organizational responsibilities as distributed, ambient and variably performed, this post-critical perspective (see also Jensen and Winthereik 2012, Yarrow and Venkatesan 2012) also indicates one-way in which management might be said to have broken up.

Critiques of Management, Standardization, and Development

Michael Power developed his critique of audit in the aftermath of the wave of New Public Management (NPM) initiatives, which massively influenced the management of public institutions in the 1990s. Not least, New Public Management became (in-)famous for its effort to *optimize* such institutions. The threat of rising public expenses had to be curtailed without citizens, now redefined as customers (du Gay and Salaman 1992) experiencing decreases in service. Creating transparency was meant to identify superfluous modes of operation and to support streamlining of the entire sector. Thus, the implementation of new forms of documentation was central to

² We conducted this fieldwork in 2011-12 as part of a more comprehensive study of partnership, infrastructures and monitoring in development aid (Jensen and Winthereik 2013).

these reforms. Across the board, organizations introduced new monitoring technologies, rendering an ever-broadening array of activities amenable to audit.

In a general sense, New Public Management followed the long shadow cast by Frederick Taylor, whose time-motion studies had been the crucial technique of early industrial optimization. NPM deployed a more differentiated set of techniques, yet it remained congenially fascinated with determining the optimal way for organizations to work (Tolsby 2000). Moreover, NPM also followed the Taylorist division of labour, according to which it was up to managers to identify the optimal *modus operandi*, while it was the task of employers to practically execute it.

One of the most enduring social scientific critiques of audit is that monitoring and evaluation systems are reductive. Only able to measure what they were designed to do, these systems are inherently narrow and inflexible, and so incapable of making visible anything different, new, or surprising. Hence, the audit systems that surround us have the primary effect of simplifying a complex reality. Countering these reductive effects, social scientists position themselves as watchdogs, vigilant voices speaking truth to power (e.g. Shore and Wright 2015).

Broadly similar critiques were developed within the anthropology of development. Drawing on Michel Foucault, James Ferguson's (1990) path-breaking study described international development as a machine for turning political problems into technical fixes. Ferguson described his own endeavour as diagnostic, but many subsequent studies were explicitly critical. These include Arturo Escobar's (1995) *Encountering Development*, which inspired a series of 'post-development' studies (e.g. Rahnema and Bawtree 1997), all of which centred on critiquing the power dynamics inherent in any development effort.

From ethnographic locations in developing countries, situated amongst those development programs intended to 'develop,' anthropologists had rich opportunities for documenting the deployment of reductive plans and projects that made little local sense, and often led to harmful consequences. Yet, this position also made it easy to depict development professionals as 'others;' at once powerful and ignorant. Indeed, Ferguson (1997, 165) described such depictions as an 'anthropological specialty.'

Our position amongst *NatureAid* professionals made this view of development professionals untenable. We began from an office in Copenhagen, which functioned as our central field site, and had only occasional opportunity to travel with our informants to *their* "field sites," in other countries. Elsewhere (Jensen and Winthereik

2013), we have described our approach as an *inverse development anthropology*, since, compared with the critical anthropology of development, which is usually premised on fieldwork located in places we only briefly visited, the sequence and the direction of the ethnography was reversed. Our point, however, is not *analytical* reversal aiming to vindicate development against its critics. Instead, we are interested in further differentiating it as a series of variegated practices (Venkatesan and Yarrow 2012). In this paper, we do so by outlining a post-critical alternative to monitoring and standardization.

Among other things, alternative hinges on the relation between performativity as exhibited in organizational optimization efforts and the performativity of social science. To situate the argument, we therefore briefly outline the key differences between our approach and some recent interpretations of performativity in the critical management literature.

Performativity and Organisational Experimentation

In *The Postmodern Condition* (1984), the philosopher Jean-Francois Lyotard argued that in the age of information ‘proofs’ had become performative. ‘The technical criterion, introduced on a massive scale into scientific knowledge,’ he wrote, ‘cannot fail to influence the truth criterion’ (1984, 46). Among other things, technical evaluations were becoming self-referential. As Michael Power would later argue, one consequence was a generally diminishing ability of evaluations to produce substantial knowledge. This decline correlated with an increasing capacity for producing self-affirming kinds of information. In more general terms, what was described was a collapse of content into form. Anthropologists of audit and standardization drew broadly similar conclusions. For example, Peter Pels argued that audit ‘actively constructs the environments in which it operates, making it more “auditable” at the same time that it renders itself thereby invulnerable to its own failure’ (2000, 142, referring to Power (1994, 7-8)).

Within management studies, some critical scholars have focused on the fact that performativity, in Lyotard’s sense, made knowledge part of ‘means-ends calculations’ (Fournier and Grey 2000, 17). To counter this tendency, they promoted an ‘anti-performative’ approach, critical of deploying instrumental reason in service of economic efficiency. Countering this approach, André Spicer *et al* (2009, 538) argued instead for ‘critical performativity,’ defined as the ‘active and subversive

intervention into managerial discourses and practices.’ Aiming to move beyond the ‘cynicism that pervades CMS [critical management studies],’ this approach would remain critical by ‘radically questioning widely accepted assumptions’ while also remaining performative by opening ‘up new ways of understanding and engaging with the discourse with the ambition to have some effects on practice’ (Alvesson and Spicer 2012, 376). Critical performativity thus assumes the possibility of directing performativity away from economic efficiency and instrumental reason and towards worthwhile socio-political aims.

This position, however, has also come in for critique. Radical aspirations aside, some have characterized it as offering a benign form of managerialism. Others have argued that it embeds overly ‘optimistic assumptions about the power of language to change certain structural realities as well as the capabilities of CMS scholars to perform emancipatory change through discourse and micro-level engagement’ (Fleming and Banerjee 2016, 263).

Yet others pointed to a more general problem with the understanding of performativity. Thus, Cabantous *et al* (2016, 201) followed Judith Butler (e.g. 2010) in arguing that one cannot *make* discourses more or less performative because performativity is a general condition of discourse. Spicer *et al* (2016) vigorously resisted this critique, depicting it as an ivory-tower preoccupation with the purity of theoretical definitions.

Even so, the critique stands. If, as has been convincingly demonstrated (e.g. Cooren 2010, Derrida 1988, Fish 1989), performativity is a condition of discourse, it is not a matter of volition. Integral to processes of material transformation (e.g. Callon 2010, Jensen 2005, Pickering 1995), one cannot, similarly, elect to be materially performative (Jensen 2007). Whether operating through discourse, materiality, or mixtures of the two, there is no way of ensuring a match between intention and performative outcome. Accordingly, the firm declaration of performative intent makes one neither more nor less performative than those who with equal determination argue against performativity. Everyone is, performatively speaking, in the same boat.

This insight is basic to our argument that optimization in *NatureAid* must be understood in a post-critical register. Since inability to dictate (performative) outcomes is a general organizational and analytical condition, the question of how to develop an improved standard within *NatureAid*, which hold our attention in the following, gains an *experimental dimension*.

In fact, an emphasis on experiments unfolding in practice is deeply embedded within science and technology studies. Thus, studies have demonstrated the multiple and variable consequences of standards and classifications (e.g. Almklov 2008; Bowker and Star 2000, Jensen 2010, Zuiderent-Jerak 2015). Doing so, they challenged the idea that technologies of optimization, like the monitoring and evaluation tools that hold our interest below, are simply 'technologies of control.' Rather than assuming the efficacy of such technologies -- with the consequence that the relevant scholarly response is always to "critique" or "radically question" them -- these studies have demonstrated that the introduction of technologies and techniques tend to introduce new margins of indeterminacy and unpredictability, and that they often generate unforeseen practical and material effects. Along these lines, we view discussions about a new monitoring and evaluation system in *NatureAid* as an *experiment in optimization*.

In our view, the emphasis on 'experiments in' rather than 'tools for' optimization is important in different ways. First, it sheds a different light on activities often rendered by critical management theory as forms of resistance. Rather than critiquing optimization *in toto*, viewing it as empirical forms of organizational experimentation makes it possible to distinguish between different forms and consequences. While some forms of optimization may indeed be resisted, responses may also include testing the capacity of standards to lead to improvements, making different monitoring and evaluation tools, or attempting to redefine what counts as optimal. In the following, we explore such forms of experimental optimization in the Copenhagen office of *NatureAid*.

Terms of Reference

Our fieldwork at *NatureAid* explored the role of monitoring and evaluation standards in establishing partnership relations in development aid. Shortly before our arrival, the Danish office had received a large grant from *Danida*, the Danish development agency. Because of the substantial amount of money involved, the development aid team within *NatureAid* was particularly attentive to *Danida*'s requirements for transparency. This attention was further amplified, since these requirements were different from those of the private funders with whom they already worked. On top of these demands, the office had to live up to the audit requirements of *NatureAid*'s own headquarters.

Within this general context, we were able to study a variety of standards: bi-annual and annual reports, the so-called ‘evaluation wheel’ used by the international headquarter, and different documents required by *Danida* and private donors. We also observed employees during their daily work, and we conducted unplanned and structured interviews with employees at different organisational levels. Finally, Brit joined a 12-day visit to monitor an ecotourism project in Vietnam, carried out in partnership with the organization’s Hanoi office (Jensen and Winthereik 2015).

Within the office, monitoring and evaluation practices were quite hard to pin down. As a set of activities that often blended into other kinds of work, such as reflecting on competences, capacities and outcomes, monitoring and evaluation had in some sense become ‘second nature’ for the team members. At the same time, however, almost every meeting generated lively discussion about the quality of the standards used for audit. Rather than focusing on a particular ‘project’ that aimed to improve audit, we try in the following to capture a sense of its organizational ambience.

Our ability to engage with a small team of environmental consultants was the outcome of a half-year of preliminary negotiations with management at the Copenhagen office. At the end of this lengthy process, we had been requested to formulate a so-called TOR, a *Terms of Reference*, describing how our study would contribute to the monitoring and evaluation efforts taking place in the NGO. Directed to the board, the final TOR stated that we would evaluate organizational efforts to evaluate, and that we would share our analyses in two workshops. Whereas our own interest was ethnographic and analytical, its pursuance was thus made possible by promises to contribute to internal organizational learning. One might therefore say that our research ended up somewhere in-between analyses *of* and *for* organizations (Jensen and Winthereik 2013, 31-51; Neyland 2008), and thus also subject to the accusation of ‘benign managerialism’ mentioned above. For reasons discussed below, however, we find this kind of critique to be rather beside the point.

In some ways our position was comparable to the position in which the evaluators that we studied found themselves. Like the conditions that framed our study hinged on negotiations across the NGO, and similar to the way in which our acceptance also depended on a mixed set of expectations, the environmental consultants were also negotiating *their* terms of reference across a number of different locations and with varied stakeholders.

Saving the World with Reports

NatureAid is located in one of Copenhagen's shabbier neighbourhoods. Once per week we travelled there by bus or bike to conduct fieldwork. Over a period of nine months we did participant observation, joined meetings and conducted *in situ* interviews with a view to understanding organizational practices of monitoring and evaluation. Often, we would 'shadow' a member of the 'development team' (Danish: *U-landsteamet*). This team consisted of four consultants who we call Eric, Magnus, Kirsten and Hanne. While *Danida* sponsored most activities in this period, the team referred directly to the local branch manager, a well-known figure who often comments on environmental issues in the Danish media. Due to *Danida's* focus on poverty reduction, the team was involved in both environmental conservation and development.

Nature Aid, we were told, is a 'knowledge organization.' This meant that the local branches are involved in creating knowledge about different aspects of environmental hazards, which is disseminated through the *NatureAid* network. As a knowledge organization, the design and implementation of projects, as well as lobbying and advocacy, all had to be based on a 'solid, preferably evidence-based' foundation (interview Eric, *NatureAid* 2010). Though the international *NatureAid* headquarter employs its own scientists to create this foundation, it also relies heavily on external consultants, commissioned to research certain topics and areas, with a view to subsequent development of benchmarks for monitoring and evaluation.

These forms of evidence were used by team members in order to lobby large corporations. For example, they might present expert evidence on the harmful consequences of palm oil production in order to convince supermarket chains to buy more environmentally friendly products. Team members also used the knowledge base to identify sites of suspect activity, like illegal logging, in their partner countries. Joining consultants in these countries on field trips, they aim to liaise with local people in order to raise awareness of environmental issues and, ideally, develop shared responses.

In general, people working for *NatureAid* were very busy handling their project portfolios, applying for funds, lobbying, travelling, budgeting -- and monitoring. They typically had too little time, and too much to do. Unhappiness with the available information and the urge to change the situation certainly also played a

role in the decision to let us study their work. At the time of our arrival, a sense that the numerous incongruent documentation requirements of headquarters and the various donors could be handled more effectively, and the feeling that current monitoring and evaluation efforts did not lead to any noticeable in-house learning, had led to the idea of developing a new standard, or, rather, re-inventing an old one.

In an interview with the branch manager, we learned that monitoring and evaluation of the overall *NatureAid* network is based on Key Performance Indicators (KPIs). Yet team members observed that the network KPIs did not produce information relevant to the variety of funders to whom they had to report. Moreover, they complained that the KPIs were useless for local processes of learning and reflection. What they sought was information they could turn into practically useful *knowledge*.

To understand the following discussions around standards for monitoring and evaluation in *NatureAid* it is relevant to note that the professionals in the development team were all, in different ways, out to 'save the world.' Several interviews highlighted the idea of 'making a difference' as a crucial motivation for working in the NGO. For example, Magnus told us that he cherished his organization's role as 'watchdog.' Contrary to *Danida*, which cannot avoid being entangled in 'all sorts of political agendas,' he felt that he was able to speak freely, thus 'raising the bar' of expectations (interview Magnus, *NatureAid* 2010).

Yet, in spite of hard work and good intentions, it was often difficult for team members to know whether the projects they ran had *really* changed anything for the better. Hence the strong interest in figuring out how their efforts might be optimized.

The employees took the obligatory monitoring and evaluation quite seriously. The reports were obligatory in any case, but the team also assumed that they *did* contribute to understanding aid efficiency at an aggregate level. At the same time, however, these efforts were largely seen as something one needed to get over with. Since the evaluations were meant for consumption and use in *Danida* or in *NatureAid* headquarters, they were not considered of much use locally.

However, this did not mean that monitoring was perceived as irrelevant in general. Hanne from the development aid team explained the situation as follows:

To a significant degree we are monitoring for ourselves. If we only monitored in order to be on good terms with *Danida*, then we wouldn't... those 5 page

status reports about something you have worked with for many years, it can easily be done while also focusing on other tasks. So we do it for our own sake. Because we would like to make some super cool projects. I really want to make a difference (Interview *NatureAid*, 2011)

However, rather than the content of submitted reports, Hanne and others insisted that what mattered was shared reflections prior to or following the submissions. In turn, the importance of these discussions faded in comparison with the continuous informal work to make and maintain relations between team members, and with project partners in other countries. Thus, Hanne explained:

I am not getting information about how a project is running through standardized formats. It doesn't happen when I receive my progress report. It happens through on-going dialogue and through project visits. The reports are fine as a backup, but these standards that tell you things are going like this and this with regards to input 1, no. What is substantial is where you get knowledge about what you actually are trying to accomplish, using taxpayers' money. What kind of difference are we trying to make? I don't get that from the documents but from other sources (Interview *NatureAid*, 2011)

Whereas 'standardized formats' were thus used to demonstrate the importance of projects to *others*, the knowledge relevant to Hanne and her colleagues had other sources, as well as other means of distribution.

The 'Evaluation Wheel'

During a mail correspondence, Hanne sent us a number of non-mandatory guidelines for improving monitoring and evaluation circulated by *NatureAid* headquarters. Among these were an 'evaluation wheel' which, depicting projects as 'cycles,' covered everything from the first project description to the final evaluation. In the image of the wheel, initial project plans and project outcomes would be joined together. If they didn't, one could embark on another round of evaluation. Among many other things, the wheel requested information about 'general practice and assumptions,' 'definitions,' 'design,' 'implementation,' 'analysis and adjustment,'

and ‘sharing.’ The wheel would also elicit information about how to ‘give and receive feedback, conduct *evaluations and audits*, and promote a culture of learning.’

To promote a culture of learning, this standard both loosened and went considerably beyond the KPIs, encouraging informal exchanges over e-mail and telephone, and emphasizing that learning is a two-way process. Thus, the evaluation wheel testified to the organizational co-existence of formal, quantitative monitoring and evaluation and qualitative, process-oriented and narrative methods. Indeed, we are in the vicinity of a view of knowledge *making* not radically different from what one encounters in qualitative sociology, socio-cultural anthropology, or critical management studies. Highlighting the important of informal communication, for example, the standard aligns with social scientific analyses, which have focused on nurturing communities of practice (Lave and Wenger (1991), see also Vann and Bowker (2001)).

In principle, the evaluation wheel offered an excellent tool for eliciting and sharing information about projects and results. We initially assumed that it would be important for the team, not least because it encouraged the simultaneous use of qualitative and quantitative methods. In reality, though, we never saw the standard used, or even mentioned, other than in the email conversation and in an interview with the branch manager about the KPIs.

Eventually we asked Kirsten whether and how the international standards were used in daily work. In response, she flatly denied *any* practical utility. ‘Fundamentally,’ she replied, ‘those standards are far too general and unspecific. There is no relation between their demands and what it makes sense for us to evaluate.’ Possibly, the standard was good at measuring certain things, but they were not the kinds of things she wanted to know:

They can be used to collect experiences about what can be measured. An increase in rhinos. They are of no use in showing us changed patterns of activity among those who cause the problems for the natural resources.
(Interview *NatureAid*, 2011)

What Kirsten opposed was not the ambition embedded in the evaluation wheel to integrate qualitative and quantitative understandings, but rather the conservationist logic underlying both the KPIs and the wheel. While good at counting rhinos, for

example, the wheel provided no tools for understanding changes in peoples' orientations to their environments. Nor did it generate information about the relation between environmental protection and poverty reduction, central to *NatureAid's* portfolio of projects funded by *Danida*. Indeed, Kirsten repeatedly contrasted the broad vision of the local branch compared with the narrow emphasis on conservation promoted by headquarters.

In her view, the evaluation wheel was simultaneously too detailed, as it required team members to spend time making many irrelevant things visible, and useless for capturing the aspects of complex project realities that truly mattered. Rather than opening up for discussions about the difficulties of balancing conservation and development work, the wheel muted these conversations.

This dissatisfaction was generally shared within the team. During our fieldwork, it led to a group decision to try to reintroduce a set of internal evaluation tools, which had been lost from organizational memory in connection with cutbacks and lay-offs a few years earlier. In Kirsten's view, remaking this standard would help the team to 'evaluate the relevance, use value, and sustainability of development work.' This, she thought, marked a fundamental difference from the kinds of monitoring and evaluation required by external stakeholders.

Minimalism was central to the plans for reinventing the homemade standard. In stark contrast with the proposal for comprehensive documentation outlined in evaluation wheel, the initial idea centred on a brief two-page list containing open-ended questions about project progress.³ Rather than requiring individual responses, the team would simply meet on a regular basis to discuss the particular questions that made most sense at the time. The local standard would institutionalize 'informal feedback,' as recommended by the international standard, but it would do so without adding extra layers of documentation and bureaucracy. Kirsten, one of the few employees who had worked in *NatureAid* when the system was last used, described its advantages as follows:

It was a good system because it created communication. It forced us to formulate some clear goals, and to collaborate across projects. It created a

³ See Mol and Law (2002: 13-14) on lists and complex knowledge.

good working environment, and there was a real urge to discuss. It was a happy period (Interview *NatureAid*, 2011)

The discussions around re-inventing the standard revolved around how to create a shared room for discussions of content and work progress. Similar to Teun Zuiderent-Jerak's (2015) argument that standardization is itself a situated practice, the homemade standard can be seen as an outcome of the differing perspectives of what constitutes work optimization held by differently situated managers and consultants. Kept deliberately vague to give new kinds of communication a chance to emerge, the consultants agreed that the new standard should not be a tool to 'bridge the gap' between managers and practitioners, but a tool to optimize *their* day-to-day operations at the office and in the field. It was a solution to a transparency problem that was opaque to the gaze of the managers – locally as well as at headquarters.

Critical Potentials

Above we have described discussions unfolding among the team about the shape of a new monitoring and evaluation system meant to be locally meaningful.⁴ Given the packed schedules of the consultants, it was far from obvious that their limited time should be spent debating the form of yet another standard. Yet, they described alternative evaluation procedures, and went on to try to convince management that such a system would have a legitimate organizational role. In fact, this effort eventually collided with a lack of managerial support. Yet, their persistence strikes us as significant. Accordingly, we use it as an entry point for appraising the potentials of the conventional critiques of management, standardization, and development.

It might be argued that the team had simply internalized the logic of the audit paradigm. After all, even as they craved self-reflection and learning, the interpretive frame remained monitoring and evaluation. Spun into the audit universe of discourse, they were only able to conceive of organizational improvement in terms of yet more technical fixes. According to this narrative, the team's interest in re-inventing the standard illustrates an invisible, yet fundamentally repressive relation (e.g. Knights and Wilmott 1989).

This story line exemplifies the 'anti-performative' critique of instrumental reason. Monitoring and evaluation is seen as performing its own universe of

⁴ See Winthereik (2003) on local experimentation with standardization in health care settings.

operation. Accordingly, the system as a whole can hardly avoid closing in on itself, becoming blind to its own blind spots. As the performative effects of evaluation become invisible, the paradigm itself becomes increasingly invulnerable. Thus, any problem will simply lead to more calls for more comprehensive evaluation.

If we consider our ethnographic material, however, nothing indicates that monitoring and evaluation was strictly closed, not to mention invulnerable. For one thing, it was not the power of standards that frustrated our informants, but rather their mixture of hubris and irrelevance. What prompted their attempts to do evaluation differently was the inability of the existing standard to do much of anything. Furthermore, team members certainly did not uncritically accept audit demands. But the problems with these demands were not that they were repressive but rather that they were uninteresting or irrelevant.

It is thus difficult to maintain that the employees of *NatureAid* were either subjugated by, or entirely ignorant with respect to, the evaluation paradigm. In fact, team members were very knowledgeable about the inability of standards to capture complex realities, and they knew well that standards are only able to make visible what they have been designed to ‘perceive.’ Rather than a clean distinction between managers who govern by numbers and resistant employees, we are witness to a situation characterized by different, and only partly related, performances of optimization.

The international standard found no practical use in *Nature Aid’s* Danish office because it was too detailed. Our informants wondered about the purpose of a standard so expansive that it allowed for no *delimitation* of context. Conversely, when they sought to re-invent their own standard, it was because they were looking for ways of knowing their practice differently. While they had no interest in encompassing forms of standardization that aimed to make *everything* subject to evaluation, they were very keen on achieving specific forms of complexity reduction by slicing reality into more relevant bits.

Here, the question of monitoring and evaluation connects with that of power. After all, the interest in making a local standard can easily be seen as an attempt by team members to wrest some control from their managers. Yet, since the new standard would also reduce complex realities, it might be a way for the team members to strengthen *their* control over external partners. Exemplifying the post-development

critique, the re-invented standard would thus exhibit the same information politics as the one various funders applied against team members, just on a smaller scale.

However, our ethnography does not show much supporting evidence for this interpretation either. Indeed, one of the recurrent complaints about the existing standards was their excessive emphasis on control rather than partnership and adaptive management. Further, the discussions about the alternative standard revolved neither around a withdrawal from existing power relations nor around the need for tightening the grip on partners.

While we are certainly not obliged to share our informants' faith in the ability of the new standard to create an open, reflexive space of learning, we are required to grapple with their experimental effort to create such a space. This is why we engage neither in an "anti-performative" critique of instrumental reason, nor in a "critically performative" effort to radically question existing organizational discourses, or in yet another critique of development hegemony.

While there is little doubt that our informants 'subjected' themselves (Foucault 1991) -- both to concrete standards and to more abstract ideals of standardization -- much hinges on what we take those terms to mean. Clearly, team members did not simply bow to the demands of standards. However, they *were* subjected in quite a different sense. They actively submitted to the demands of the new standard, by *conferring to it the power to delimit the scope of their work*. They did so in the hope that such subjection might open up new possibilities for action (Gomart and Hennion 1999).

Good Standards Constrain Reality

What makes a standard for evaluation *good*? Based on our ethnography, it appears that a good standard should not seek to exercise more control over organizational routines. Nor should it seek encapsulate all dimensions of development projects. After all, the entry point for remaking the standard was that the evaluation wheel, while claiming to be comprehensive, failed to grasp much of what mattered in practice. Given these experiences, the good evaluation standard should be tailored to support a different sets of goals and capable of capturing specific bits of development work differently. At issue would not be a competition among standards, but simply co-existence (which, of course, can take many forms).

As this also makes clear, a good standard does not try to represent the world in its complexity. Quite to the contrary its point is to prune reality, like a gardener prunes a wildly growing bush. A good standard needs to make present just the kinds of things team members care about while temporarily removing what is irrelevant, such as the obviously different contexts of projects dealing with sustainable logging in East Africa or shrimp farms in the Greater Mekong. This is why the hope team members had for the new standard can be described with Emilie Gomart's (2004) notion of 'generous constraints.' Our informants were quite willing to submit to standards if their constraints were generous, for they were *looking for* constraints that would make it possible to learn together and be subjected to relevant forms of collegial scrutiny.

This suggests that conventional critiques directed against the reductions of standards move too quickly. The notion that reduction is inherently negative misses the central point that *particular* reductions are the explicit and, indeed, often *reflexive* purpose of standards. Rather than lamenting the reduction of complexity, we are directed towards exploration of the specific constraints standards put in place.

It no longer surprises management theorists, sociologists, or anthropologists that formal systems and standards entail de-contextualization. Everyone knows that local interpretations and patterns of action, underhanded political negotiations and invisible economical incentives disappear when translated into the factual language of monitoring and evaluation. However *NatureAid* team members are as aware of this as academic critics. They know very well that what is demanded of standards – demands like ensuring coherence in projects spanning donors in Copenhagen and villages in Malawi or Vietnam -- is very difficult, if not totally unachievable, since the contexts are almost incommensurable (Rottenburg 2009). But this is the *starting point* of their experiment in optimization. It is *because* the team cares as much as critical social scientists about the varied, uncontrollable effects of standards, or even more, since their job and their passion depend on them, that they invest time in imagining new forms of evaluation.

Reductions and Additions

There is no general difference between “anti-performative” positions that aim to resist instrumental reason and “critically performative” ones that aim to subvert practices from within. Critical theories of management, too, operate performatively, *as part of*

reality. Even as they denounce the reductions of management discourse, instrumental reason, standards or optimization, they are also reducing reality in their own ways. Having access to no all-seeing, non-reductive, meta-position, critics are part of the same game as the managers and organizations they write about. Thus, critique loses its epistemic privilege, without gaining any specific performative efficacy.

Yet even as critical analyses and audit tools both reduce the world they also both add to it. New standards and indicators do new things to organizations, just as new research may change academic discussions. Hence, it becomes possible in both cases to query the relevance and generosity of the constraints and reductions imposed, with a view to understanding their additive effects. In this post-critical landscape, critique thus gives way to experimental engagements with particular bits of the world they would like to understand, or see changed.

From this vantage point, the problem with managerial efforts to monitor and evaluate their employees and projects, even in their informal or otherwise ‘soft’ dimensions, is not that they reduce lived experience or organizational complexity to an image of instrumental reason. Rather, the problem has to do with the particular way in which audit adds to organizational reality. Drawing on Marilyn Strathern (2000a), this addition can be characterized as *self-propelling*.

Once audit information has been produced, it begins to generate calls for action. Moreover, irrespective of the organizational “value” (however measured) of audit information, it tends to be very difficult to get rid of. For Strathern (2000b), these interrelated characteristics constitute the tyranny of transparency. We can see the re-invention of the *NatureAid* standard as an experimental attempt to ward off, or at least minimize, this tyranny.

However, a residual problem remains with this diagnosis. As other critical discussions of audit, it presents transparency as a homogeneous phenomenon. Yet, as we have seen, even within the limited organization and professional domain of *NatureAid*, transparency is a variable phenomenon. It is understood differently, and it is embedded within different practices and standards. When critics speak of transparency as a homogeneous form, they are overlooking the multiplicity of its empirical forms. Further, while any effort to create transparency involves the imposition of a particular kind of order on the world, it simultaneously generates its own ‘other’ -- constituted by everything not rendered transparent. Thus, transparencies produced by KPIs or by the evaluation wheel are different from those

imagined or produced by the homemade standard. Rather than transparency in the singular, there are many relative transparencies.

Even within *NatureAid*'s Danish office, optimization is variably enacted, shaped by diverse, competing visions, motivations, strategies and standards. When different forms of optimization and transparency bump into one another, they may give rise to frustrations, controversies and critique. They may create dents in the relations between branch offices and international headquarters. What appears like an obvious strategy for creating organizational transparency from one position may seem opaque or even incomprehensible from another.

Learning from Experiments in Optimization

The organizational discussions about how to best monitor and evaluate development work in *NatureAid* offer a platform for considering the analytical purchase of critical perspectives on audit and optimization. Whereas such perspectives draw the conclusion that standards reduce complex realities to instrumental reason, we have argued that this distinction itself breaks down, both in practice and in theory.

In 'The Critiques of Utility,' the literary scholar Barbara Herrnstein Smith vividly describes the basic dualist assumption informing critical social science:

the force of the opposition/segregation of the discourses of value is most evident, perhaps, in the recurrent struggles between two kinds of calculation or cost-benefit analysis: on the one hand, the kind, so named, that frames its objective as the efficient arrival at a specific and readily identifiable ... "bottom-line" and, accordingly, ignores or downplays less readily measurable and less comparable costs, risks and benefits... and, on the other hand, and typically in antagonistic relation to the first kind, *another* calculation, *not named as such*, that characteristically foregrounds and promotes exactly what was ignored by the first (Smith 1988, 133, emphasis in original)

Critical management studies, as well as sociological critiques of standardization and anthropological critiques of development, are all situated within, and contributing to, Smith's second discourse of value, the 'other calculation.' Focusing on the 'subtle, diffuse...and heterogeneous' (Smith 1988, 133) aspects of social and organizational

life, these studies take audit and standardization *in general* to task for reducing just these complex dimensions.

Describing *both* discourses as calculative, however, Smith emphasized that ‘the categories and considerations with which they deal and the operations they perform are only *relatively* and *locally* distinguishable from one another’ (Smith 1988, 133). As we have indicated, monitoring and evaluation at *NatureAid* embeds different calculations of value (Helgesson and Muniesa 2013) that are also only relatively and locally distinguishable.

This observation continues to hold once we move from the empirical setting to that of academic analysis. For the difference between development consultants and social scientists, or between ‘field’ and ‘desk,’ is itself less clear-cut than critical scholars may like to imagine.

Thus, Marilyn Strathern notes that

The auditor or assessor ... is little different from the anthropologist translating across cultures. The anthropologist’s analytical categories (social structure, cultural values, modes of organisation) turn one kind of description (e.g. peoples’ perceptions of themselves) into another and thereby conceal certain truths in revealing others (Strathern 2000b, 70)

More starkly, Frederick Cooper and Randall Packard’s (1997, 26) observed that social scientific concerns ‘with context and complexity is neither more nor less separable from a self-serving professionalism than the development practitioner’s concern with the replicability of project design.’ While critical management theorists and anthropologists have their eyes trained on the ‘self-serving’ reductions of consultants and development professionals, Cooper and Packard’s point is that their own emphasis on context and complexity is equally subservient to the academic networks and discourses of which the critics are part.

Cooper or Packard are not, of course, arguing for the *irrelevance* of context and complexity. They are simply noting that, being ‘neither more nor less separable’ from their contexts than the positions and practices of development consultants, academic diagnoses are never made from an elevated meta-position. Instead, they are also performative elements that operate within specific networks. Accordingly, the critical perspectives of management scholars are not inherently superior to those of

their informants – for example those of *NatureAid* professionals. Yet, given these professionals were in fact trying to take simultaneous account of ‘complexity and replication,’ we have suggested that organizational researchers might learn something from their experimental attempts.

As part of their efforts to create and maintain relations with partners as diverse as *Danida* and villagers in partner countries, *NatureAid* team members wondered about how to relevantly optimize their own practices. They constantly tried to improve things. They attempted to adjust formal evaluations to make them just a bit more useful. They tested new modes of collaborating with their partners (Jensen and Winthereik 2015). And, as we have described, they experimented with developing a standard that would impose generous constraints.

Team members did all these things not because they were blinded by audit, or were subjugated by it, but rather because they were all too aware of the world’s complexity. They urgently needed to delimit this complexity, or their projects would collapse as quickly as a house of cards in the wind.

Here we can make a distinction between optimization as a normative demand and as an experimental disposition. While optimization of organizational performance, including self-management, was obviously operative in *NatureAid* as a normative demand, the practical achievements of the organization nevertheless crucially depended on employees’ concrete experiments with optimization.

When team members collaborate with new partners, the situation is inherently experimental, since they cannot know in advance what is the best course of action. Indeed, they are often quite unsure of whether and how their precarious interactions with people they do not know well may lead to successful projects. Invariably, these optimization efforts thus entail the risk of navigating into unknown waters. The risk is real, for projects can easily go awry, and foreign agendas can translate the aspirations of the employees beyond recognition. This is why experimental optimization efforts are directly tied together with the idealist passion for ‘saving the world’ that forms a backdrop to work in *NatureAid*.

Focusing on experiments in optimization within *NatureAid* has facilitated a re-description of audit as a varied set of practices and aspirations, embedded in standards that generate relative transparencies. Elucidating the gaps and mutual opacities between forms of monitoring and evaluation, we have further emphasized that management itself is heterogeneous. Rather than a dichotomy between controlling

managers guided by instrumental reason and resistant professionals aiming to elude control, both groups are involved in optimizing efforts. These efforts, however, deploy different standards, and have somewhat different (but not totally incompatible) aims.

Our study thus exemplifies one way in which management has ‘broken up.’ The point is neither that management shows signs up breaking *down*, nor that it threatens to collapse under the onslaught of critique. Rather, management in *NatureAid* appears ‘broken up,’ in that it is a distributed, ambient activity, variably performed by different actors who use a cohort of standards and audit tools. This fragmentation suggests that management in practice does not depend on the kind of closed, bulletproof system beloved as a target of critique by critical management studies. To the contrary, it is because organizational members recognize that the system is potholed, full of gaps, inconsistencies, and indeterminacies that they engage in experiments in optimization. Their work can be seen as a practical, experimental version of what Barbara Herrnstein Smith (1988) called ‘alternative perspectives for critical theory.’

Here, we have therefore tried to create some analytical elbowroom for learning from our informants’ alternative perspectives what constitutes good forms of optimization. In our view, management and organizational researchers would do well to heed their experimental disposition.

References

Almklov, Petter. 2008. “Standardized Data and Singular Situations.” *Social Studies of Science* 38(6): 873-97.

Alvesson, Mats, and André Spicer. 2012. “Critical Leadership Studies: The Case for Critical Performativity.” *Human Relations* 65(3): 367-90.

Bowker, Geoffrey C. and Susan Leigh Star. 2000. *Sorting Things Out: Classification and its Consequences*. Cambridge, MA & London: MIT Press.

Butler, Judith. 2010. “Performative Agency.” *Journal of Cultural Economy* 3(2): 147-61.

Cabantous, Laure, Jean-Pascal Gond, Nancy Harding, and Mark Learmonth. 2016. “Critical

Essay: Reconsidering Critical Performativity.” *Human Relations* 69(2): 197-213.

Callon, Michel. 2010. “Performativity, Misfires and Politics.” *Journal of Cultural Economy* 3(2): 163-9.

Cooper, Frederick, and Randall M. Packard. 1997. *International Development Aid and the Social Sciences*. Berkeley, Los Angeles & London: University of California Press.

Cooren, Francois. 2001. *Action and Agency in Dialogue: Passion, Incarnation and Ventriloquism*. Amsterdam & Philadelphia, PA: John Benjamins Publishing.

Derrida, Jacques. 1988. *Limited Inc*. Evanston, IL: Northwestern University Press.

du Gay, Paul, and Graeme Salaman. 1992. “The Cult(ure) of the Customer.” *Journal of Management Studies* 29 (5): 615-33.

Escobar, Arturo. 1995. *Encountering Development: The Making and Unmaking of the Third World*. Princeton, NJ: Princeton University Press.

Ferguson, James. 1990. *The Anti-Politics Machine: "Development", Depoliticization, and Bureaucratic Power in Lesotho*. Cambridge: Cambridge University Press.

Ferguson, James. 1997. “Anthropology and Its Evil Twin: ‘Development’ in the Constitution of a Discipline.” In *International Development and the Social Sciences: Essays on the History and Politics of Knowledge*, edited by Frederick Cooper and Randall M. Packard, 150-76. Berkeley: University of California Press.

Fish, Stanley. 1989. *Doing What Comes Naturally: Change, Rhetoric, and the Practice of Theory in Literary and Legal Studies*. Durham, NC & London: Duke University Press.

Fleming, Peter, and Subhabrata Bobby Banerjee. 2016. “When Performativity Fails: Implications for Critical Management Studies,” *Human Relations* 69(2): 257-76.

Foucault, Michel. 1991. *Discipline and Punish: Birth of the Prison*. London: Penguin.

Fournier, Valérie and Chris Grey. 2000. "At the Critical Moment: Conditions and Prospects for Critical Management Studies." *Human Relations* 53(1): 7–32.

Gomart, Emilie. 2004. "Surprised by Methadone: In Praise of Drug Substitution Treatment." *Body and Society* 10: 85-110.

Gomart, Emilie, and Antoine Hennion. 1999. "A Sociology of Attachment: Music, Amateurs, Drug Users." In *Actor-Network Theory and After*, edited by J. Law and J. Hassard, 121-147. Oxford: Blackwell Publishers.

Helgesson Claes F., and Fabian Muniesa. 2013. "For What It's Worth: An Introduction to Valuation Studies." *Valuation Studies* 1 (1): 1-10.

Jensen, Casper Bruun. 2005. "An Experiment in Performative History: The Electronic Patient Record as a Future-Generating Device" *Social Studies of Science* 35(2): 241-267.

Jensen, Casper Bruun. 2007. "Sorting Attachments: Usefulness of STS in Health Care Policy and Practice." *Science as Culture* 16(3): 237-51.

Jensen, Casper Bruun. 2010. *Ontologies for Developing Things: Making Health Care Futures Through Technology*. Rotterdam: Sense.

Jensen, Casper Bruun and Brit Ross Winthereik (2012) "Recursive Partnerships in Global Development Aid" in Soumya Venkatesan and Thomas G. Yarrow (eds.) *Differentiating Development*. London and New York: Berghahn, pp. 84-10.

Jensen, Casper Bruun and Brit Ross Winthereik. 2013. *Monitoring Movements in Development Aid: Recursive Partnerships and Infrastructures*. Cambridge, MA: MIT Press.

Jensen, Casper Bruun and Brit Ross Winthereik. 2015. "Test-Sites: Attachments and Detachments in Community Based Eco-Tourism." In M. Candea, J. Cook., C. Trundle & T. Yarrow. *Detachment: Essays on the Limits of Relational Thinking*. Manchester: University of Manchester Press, pp. 197-218.

Knights, David and Hugh Willmott. 1989. "Power and Subjectivity at Work: From Degradation to Subjugation in Social Relations." *Sociology* 23(4): 535-558.

Latour, Bruno. 2004. "Why has Critique Run Out of Steam: From Matters of Fact to Matters of Concern." *Critical Inquiry* 30 (2): 225-48.

Lave, Jean, and Etienne Wenger. 1991. *Situated Learning: Legitimate Peripheral Participation*. Cambridge: Cambridge University Press.

Lyotard, Jean-Francois. 1984. *The Postmodern Condition: A Report on Knowledge*. Minneapolis, MN & London: University of Minnesota Press.

Mol, Annemarie, and John Law. 2002. "Complexities: An Introduction." In *Complexities: Social Studies of Knowledge Practices*, edited by J. Law and A. Mol, 1-23. Durham, NC & London: Duke University Press.

Neyland, Daniel. 2008. *Organizational Ethnography*. London: SAGE.

Pels, Peter. 2000. "The Trickster's Dilemma: Ethic and the Technologies of the Anthropological Self." In *Audit Cultures: Anthropological Studies in Accountability, Ethics, and the Academy*, edited by M. Strathern, 135-172. London & New York: Routledge.

Pickering, Andrew. 1995. *The Mangle of Practice: Time, Agency, and Science*. Chicago, IL: University of Chicago Press.

Power, Michael. 1994. *The Audit Explosion*. London: Demos. Available at <http://www.demos.co.uk>.

Power, Michael. 1997. *The Audit Society: Rituals of Verification*. Oxford: Clarendon.

Rahnema, Majid and Victoria Bawtree. 1997. *The Post-Development Reader*. London: Zed Books.

Rottenburg, Richard. 2009. *Far-Fetched Facts: A Parable of Aid*. Cambridge, MA: MIT Press.

Shore, Cris, and Susan Wright. 2015. "Audit Culture Revisited: Rankings, Ratings, and the Reassembling of Society." *Current Anthropology*. 56 (3): 421-444.

Smith, Barbara Herrnstein. 1988. *Contingencies of Value: Alternative Perspectives for Critical Theory*. Cambridge, MA: Harvard University Press.

Spicér, Andre, Mats Alvesson, and Dan Kärreman. 2009. "Critical Performativity: The Unfinished Business of Critical Management Studies," *Human Relations* 62(4): 537-60.

Spicér, Andre, Mats Alvesson, and Dan Kärreman. 2016. "Extending Critical Performativity." *Human Relations* 69(2): 225-49.

Strathern, Marilyn, ed. 2000a. *Audit Cultures: Anthropological Studies in Accountability, Ethics, and the Academy*. London & New York: Routledge.

Strathern, Marilyn. 2000b. "The Tyranny of Transparency." In *Shifting Boundaries of the Real: Making the Invisible Visible*, edited by H. Nowotny and M. Weiss, 59-78. Zürich: VDF Verlag.

Tolsby, June. 2000. "Taylorism Given a Helping Hand. How an IT System Changed Employees' Flexibility and Personal Involvement in Their Work." *Journal of Organizational Change Management* 13 (5): 482-492.

Vann, Katie and Geoffrey C. Bowker. 2001. "Instrumentalizing the Truth of Practice" *Social Epistemology* 15: 247-262.

Zuiderent-Jerak, Teun. 2015. *Situated Intervention: Sociological Experiments in Health Care*. Cambridge, MA and London. The MIT Press.

Venkatesan, Soumhya and Thomas Yarrow, editors. 2012. *Differentiating Development: Beyond an Anthropology of Critique*. Oxford: Berghahn.

Winthereik, Brit Ross. 2003. "We Fill in Our Working Understanding: On Codes, Classifications and the Production of Accurate Data." *Methods of Information in Medicine* 42: 489-496.

Winthereik, Brit Ross and Casper Bruun Jensen. 2014. "Optimeringseksperimenter: Postkritiske perspektiver på monitorering og evaluering i en miljøorganisation," *Tidsskrift for antropologi* 70: 171-190.