



Evaluation at the Endgame: Evaluating Sustainability and the SDGs by Moving Past Dominion and Institutional Capture

Andy Rowe

Abstract

Three facts underlay this chapter. First, the human system and all our ambitions for improving the human system depend on sustainable natural systems. Second, we do not have much time. On track to fall well short of all sustainability goals, the climate and sustainability crises grow and extinction looms. Third, up to this point evaluation has shown little interest in sustainability, yet evaluation potentially addresses the very questions that are central to informing and guiding rapid adaptation of human behavior to successfully surmounting extinction.

Business-as-usual evaluation will not suffice. At the endgame with extinction looming, we need an evaluation that is more nimble, keeps up with rapidly accelerating knowledge, is relentlessly use-seeking and that guides the way to joined-up approaches. The evaluation we need will systematically mainstream sustainability across all evaluations and interventions, in all evaluation criteria and standards. For this, all evaluations will always address nexus where human and natural systems join and incorporate knowledge and methods from both systems. Existing evaluation knowledge

is well suited to this task, as are knowledges in biophysical sciences. We know and promote knowledge processes for integrative evaluation and are starting to shift toward the requirements for evaluation at the nexus. As this chapter shows, the anchors holding us back are political, not technical.

Introduction

Every line of evidence leads us to conclude that the threats to sustainability of the planet and the life it supports are very real, large, multi-faceted and imminent. And yet globally we are falling well short on milestones such as the 2030 Agenda for Sustainable Development and 2050 carbon reduction goals. We have pushed natural systems beyond their capacity to adapt and continue to provide the services on which we depend. We are at the endgame on this planet.¹

With some important exceptions, evaluation globally has not recognized the overwhelming evidence that sustainability is a matter worthy of our attention. Sustainability is a materially differ-

¹In revising this chapter, I came across the work of Robert Nadeau that seems to presage some of my own work. He wrote a decade earlier about the shortcomings of neoclassical economics in not addressing the connection between environment and economics that he attributes to a two culture (economics and environment) thinking (Nadeau, 2008) identified initially by C. P. Snow (Nadeau, 2006).

A. Rowe (✉)
ARCeconomics, Maple Bay, BC, Canada
e-mail: andy@arceconomics.com

ent matter than those that evaluators are accustomed to addressing because there is a hard stop if we fall short; absent significant improvements in our performance, that hard stop is a clear pathway to extinction. Meaning that evaluation at the endgame is different from business-as-usual evaluation. As with chess, the sustainability endgame needs to be fully goal focused and must fully commit all resources to strategies to achieve checkmate.

This chapter is concerned about the character of evaluation that will enable the field to make useful contributions at the endgame. The most fundamental change is from evaluations' almost monastic focus on the human system to systematic consideration of all interventions (projects, programs, strategies, policies) in their nexus location where both human and natural systems are present, have influence, provide value, and are affected.

The underlying mechanism for this monastic, human-centered worldview lies in the rootstock of evaluation that is said to be provided by Western social and management sciences with accountability, social inquiry, and social research methods as the trunk of the tree (Christie & Alkin, 2008; Alkin, 2004). That evaluation rootstock is embedded in and draws nutrition from the accumulated soils of Judeo-Christian society strongly infused with dominion, a worldview in which humans have ascendancy over other living and nonliving things, and over other peoples (Rowe, 2018). Humans, and of course especially those of European origin (i.e., white), are at the top of the heap; all else serves. Nonhuman living and nonliving things that constitute the natural systems on which all life depends are regarded as resources to be freely extracted to support humans. And while social sciences and evaluation are adapting to recognize and address how dominion has shaped thought and practice (e.g., gender bias, racism) the presumption that only humans have value and therefore merit consideration continues virtually unchecked in evaluation.

Accountability is one of the stems of the evaluation tree effectively partitioning governance

structures from interventions at all levels so that connectivity to public policy goals is truncated (Chelimsky, 2012). It is an important mechanism for the observable, inverse relationship between public expenditures and the status and trends on conditions targeted by public policy such as public health and education (Williams, 2019). Sustainability is about connected systems while accountability is about partitioned systems, making pursuit of sustainability at odds with contemporary approaches to accountability. Accountability is an important authorizing mechanism bringing dominion into evaluation with the unintended effect of imparting a systematic positive bias to evaluation (Rowe, 2019b).

The COVID-19 pandemic provided dramatic evidence that human and natural systems are connected (Patton, 2020b). The virus reached us along pathways created by our relentless incursions into natural systems. The inverse and causal relationship between contemporary forms of economic growth and environmental health have been starkly shown with the slowing of economic and social activities causally linked to reduced incidence of some important health conditions such as asthma and reductions in GHG emissions from economic downturn. The economic downturn has resulted in falling petroleum prices, making it less expensive to produce virgin plastic from fossil fuels as compared to recycling. At the same time, demand for disposable (plastic) protective equipment has increased manyfold. For example, daily single-use plastic medical waste (gloves, masks, and gowns) in Wuhan at the peak of the pandemic there increased sixfold compared to pre-pandemic averages (Adyel, 2020), all of which is disposed in landfills. Demand for plastic packaging is estimated to have increased by 5.5%, strongly related to the increased consumption of take-out foods; plastic deposits in landfills increased by 1400 tons during the 8-week shutdown in Singapore (Adyel, 2020). That is, the pandemic reduced economic activity, decreasing demand for fossil fuels and lowering their price, leading to increased fossil fuel use to produce single-use plastic commodities. This resulted in increased deposits in landfills and in unmanaged streams of disposal, a good example

of connectivity from public health to economy to environment.

The evaluation worldview must shift to acknowledge that human life is intrinsically contingent on healthy natural systems with which we are coupled² and that we must end the unnecessary harm we cause and move to restoring critical environmental values. Indigenous worldviews are instructive; for example, we might take direction from Daniel Wildcat from Haskell Indian Nations University:

Think of how our worldview changes if we shift from thinking that we live in a world full of resources to a world where we live among relatives. (Zak, 2019)

Evaluation does not address the natural system for social and political reasons, but we have the knowledges, tools, and methods needed to renovate evaluation by drawing on a broad palate including evaluation, social and biophysical sciences, conflict resolution, law, and other fields (Patton, 2020b; Rowe, 2018). And emerging efforts by evaluators are starting to build foundations for incorporating sustainability into evaluation, such as in Blue Marble Evaluation (Patton, 2020a) and Better Evaluation (2020). The need for these efforts is amply demonstrated by two recent stocktakings showing evaluation to be only in the early stages of addressing nexus, that development evaluation appears to lead national and sectoral efforts, and that the intellectual infrastructure for nexus evaluation can only be described as weak (Sustainability Working Group, Canadian Evaluation Society [CES], 2020; United Nations Evaluation Group Working Group on Integrating Environmental and Social Impact into Evaluations [UNEG Working Group], 2020).

This chapter's focus is on evaluation at the endgame. I begin with the findings of the two sustainability stocktakings to describe where evaluation is now with respect to systematically

incorporating sustainability into evaluation—effectively our starting point for the endgame. The findings clearly point to evaluation's almost singular focus on human systems and to an intellectual infrastructure that is not fit for the purpose of incorporating sustainability. I then briefly reprise my arguments that the cause for this state of affairs lies in a worldview of dominion whereby humans, and especially white humans, hold dominion over all other living and nonliving things. This worldview is pervasive in social science and evaluation, with accountability serving as a key mechanism authorizing disregard of the natural system in evaluation. To these earlier arguments I add institutional capture as a further mechanism separating human and natural systems in evaluation and use the example of the SDGs to illustrate this. I then return to the endgame, illustrating some fundamental differences between evaluation needed for the endgame and the evaluation we have now.

Taking Stock on Evaluation Practice and Resources on Sustainability

Two recent and complementary stocktaking efforts have assessed current evaluation practice and resources to incorporate sustainability. The UNEG Working Group on Integrating Environmental and Social Impact into Evaluations completed a stocktaking of evaluation policy and guidance on social and environmental considerations and of practices of UNEG member evaluation offices in addressing social and environmental considerations (UNEG Working Group, 2020).³ The stocktaking is to contribute to deliberations about a common UN-wide approach for incorporating environmental and social considerations into all evaluations (whether or not the evaluand is an environmental program). The second stocktaking was conducted by the Sustainability

²I use the concept of *coupled systems* to refer to the dynamic complex relationships across intimately connected human and natural systems (Liu, 2007; Ostrom, 1990; Rowe, 2019b) and *nexus* to refer to evaluation of sustainable development with interlinked social, economic and environmental dimensions (Uitto, 2019).

³UNEG is a professional network that brings together the evaluation units in the UN system, including the various UN departments, specialized agencies, funds and programs (<http://unevaluation.org/>) as well as non-UN organisations such as the GEF IEO.

Working Group of the Canadian Evaluation Society (CES) for two purposes: to assess the extent to which sustainability has been addressed in federal evaluations and by other governments and organizations in Canada and by Canadian evaluators working internationally, and to assess the intellectual infrastructure for evaluating sustainability in Canada and the United States (CES, 2020). The CES stocktaking is informing consideration of how the CES can mainstream sustainability in its own work and in evaluation in Canada. The CES stocktaking report was completed in 2020 with much of the work undertaken on a pro-bono basis by four leading Canadian consulting firms.⁴

These two undertakings cover a wide swath of global evaluation with UNEG addressing development evaluation and the CES addressing evaluation at national and sub-national levels while also assessing the Canadian and U.S. intellectual infrastructure for mainstreaming sustainability. Together, these two stocktaking efforts provide powerful evidence that the evaluation field is, at best, mildly and only recently addressing sustainability and that the social dimension is the priority for evaluation.

The two stocktaking efforts clearly showed that sustainability is largely missing in action from evaluation in the UN system and in Canada, and from the intellectual infrastructure for evaluation in the United States and Canada.

- The UNEG stocktaking also revealed that, first, coverage of the social system is also only partial and, despite heightened awareness of social–natural systems interaction, evaluation guidance on environment is extremely limited; and second, that the over-arching need emerging from documentary analysis and survey responses of UNEG member agencies is for a comprehensive document providing advice on how to evaluate the interactions among social and environmental considerations within the framework of UN activities

in support of the SDGs (UNEG Working Group, 2020, p. 6).

- The CES stocktaking showed sustainability and consideration of the natural system to be largely missing from federal evaluations conducted in 2016–2018, with Global Affairs Canada being a notable exception, and that the intellectual infrastructure in Canada and the United States for evaluation in the natural system is very limited.

The Canadian stocktaking is worth highlighting given the strong and long-standing evaluation infrastructure:

- The CES is the elder national evaluation organization among its global peers, membership per capita is highest relative to peer organizations, national training programs have been in place since the mid-1990s, and the CES developed the first evaluator credentialing in 2009.
- The Canadian government enacted a government-wide measurement and evaluation system in 1977 and the National Evaluation Policy in 1994 and 2001, requiring all federal programs and initiatives of material importance (roughly greater than \$5 million CDN) to be evaluated at least once every 5 years. This ensured that all federal departments have a strong evaluation function and that supporting evaluation in their departments and responding to evaluations is an important part of the performance criteria of federal senior managers.
- Provinces and territories also have evaluation functions and requirements, as do other levels of government such as school boards and health agencies.

For evaluation function and infrastructure, Canada is a global leader. Canada also has signed most international climate and sustainability protocols and agreements and the elected government platform and positions have, since 2015, accorded sustainability and climate a strong priority.

Given the relative strength of evaluation in Canada and wide acceptance of the importance

⁴Baastel, Goss Gilroy, Prairie Research, and Universalis.

of climate and sustainability, it is reasonable to expect more positive observations than the sustainability stocktaking showed. The stocktaking had four elements:

1. A review of all federal evaluations from 2016–2018 revealed only a very tiny portion addressing nexus or sustainability. Global Affairs Canada was the leader, associated with its responsibilities for international climate and sustainability agreements. Natural resource-focused departments only evaluated human system effects; that is, departments in the Canadian government whose mandates included natural resources conducted evaluations from an extraction stance.
2. A review of Canadian philanthropic, nongovernmental, and First Nation evaluations did not identify much in the way of evaluations addressing nexus, although they did address natural systems when this was the focus of funding. Evaluations from these sectors rarely considered both human and natural systems.
3. Examination of whether Canadian-based evaluators working internationally considered the natural system and nexus did identify international examples where this occurred.
4. And perhaps most concerning, the intellectual infrastructure for nexus evaluation or even just evaluation of natural system effects is almost asymptotic to zero; that is, the natural system does not appear in peer-reviewed evaluation literature in Canada and the United States,⁵ conference presentations, gray literature, and professional and university-based training. For example, just 4% of published papers in the four leading North American evaluation journals addressed natural system matters and only a few of these addressed nexus.

The findings of the two stocktakings are sobering but also encouraging. They are sobering in their confirmation that the evaluation field has little or no presence and little existing capacity in

contributing to sustainability, the leading issue of the day. But we can find encouragement because they clearly point to a growing recognition that sustainability is a top matter and to an interest in addressing sustainability as a priority.

Given the similarity of findings of the UNEG and Canadian efforts, a search for the systematic origins for the clear prioritization in evaluation of the human over the natural system, and the separation of the two systems, is reasonable. The next section proposes that the origins lie in a dominion-infused worldview asserting that humans are imbued with rights over all else—basically colonization of the planet to serve humans. Accountability structures have served as an important mechanism framing evaluation from a dominion perspective, and global and national governance units have sought to capture the resulting siloed landscape.

Dominion, Accountability, and Institutional Capture

The two stocktaking efforts clearly show that evaluation strongly prioritizes social matters, has very limited capacity to address natural systems,⁶ and only rarely, across the vast landscape of evaluations covered by the two stocktaking efforts, are the two systems, human and natural, considered together.

I offer an explanation that evaluation rests on knowledge that itself rests on a worldview of dominion in which humans, and especially humans of European origin, have dominion over all other living and nonliving things and regard these as resources for use as humans see fit. Social inquiry and social research methods are said to be the rootstock of evaluation (Alkin, 2004), but I argue that they draw their nutrition from the terroir of dominion (Rowe, 2019b). The other rootstock of evaluation is said to be accountability. This management construct is layered on top of dominion and is the second causal force that has contributed to an almost monastic focus

⁵The stocktaking was limited to Canada and the United States for purposes of feasibility only.

⁶Natural systems are inclusive of environmental impacts as addressed by the UNEG stocktaking.

on the human system by bounding accountability, and consequently evaluation, to the intent and boundaries of interventions severing or at least loosening connection to the public policy goals for which they exist and to other efforts addressing those goals. Third, without structure that recognizes the connectivity between human and natural system goals and the dependence of the social system on the natural systems, the SDGs offer a goal structure, initially and still today, in which the natural system does not need to be considered. This section provides a brief overview of how dominion, accountability, and institutional capture contribute to evaluation's overwhelming focus on the social system and neglect of the natural system, as reflected by the UNEP stocktaking.

Dominion

Evaluating sustainability first requires systematically recognizing and addressing those elements of both the human and natural systems that influence and are influenced by the evaluand. The stocktaking efforts showed evaluation to have an overwhelming focus on the human system, reflecting a dominion worldview where humans are ascendant and all other things, living and natural, can be extracted and deployed for human use. This is an implausible position: If human life depends on what we draw from the natural system, then the natural system must have value to the human system. The position that the natural system has no value and need not be considered has deep roots in social science and economics, which in turn are rooted in Judeo Christian worldviews and associated beliefs about dominion. Dominion is quite a simple concept whose existence is undeniable but, like any deeply embedded concept, it can be challenging to recognize and address. Dominion also provides a causal connectivity between the treatment of colonized and subjugated peoples and the treatment of other species and elements in the natural system. Indeed, one of the rationales for the actions of colonizers was the superiority of their worldview over the very different worldviews of many

of the colonized peoples who regarded themselves and other living and nonliving things as equal and part of a whole.

Dominion means that other living and natural things do not have value, that they exist to serve humans, and any monetary value ascribed to them results from ownership or regulated rights that provide the ability to control access and use. A classic example of dominion in action was the construction of massive dams for electrical generation in pursuit of industrial and economic development. Early critiques and resulting modification of cost benefit and other analysis of dams recognized and evaluated the direct losses to humans above and below the dams. But only recently have the ecosystem losses from flooding above the dam and water loss below the dam begun to be imputed, although on a limited basis. Because living and natural things other than humans were not valued, no mechanism was in place to recognize their importance and scarcity, directly causing relatively unfettered extraction and destruction—the fundamental cause of the sustainability crisis and climate change.

The issue of temporal and spatial scales is another way that dominion and accountability have led to evaluation's monastic focus on the human system. Systems are by their nature coupled, extensive, and dynamic, each with a wide range of temporal and spatial scales and often very diverse units of account (Rowe, 2012). Human temporal and spatial scales differ significantly from scales relevant to the natural system, and, of course, with a dominion-infused worldview, the units of account that matters are human. When the natural system is considered, it is usually from an extraction perspective in terms of utility to humans, not as a coupled system meriting its own place in evaluations.

Evaluation is a human system activity usually conducted from temporal scales meaningful to the aspects of the human system that is commissioning and undertaking the evaluation. By their nature, effects of a human or natural intervention have broad reach, well beyond the temporal and spatial reach of the intervention. Evaluations are aligned with the programmatic schedules of interventions and usually extend backward to the

start of the intervention and forward to some programmatic or arbitrary time, usually less than 10 years from their start. These temporal scales bear no relevance for the temporal scales of natural system elements that can range from centuries to moments.

The value and function of natural systems is not the only consequence of dominion. Clearly, racism and misogyny are causally linked to the dominion of white, European-origin males. To illustrate, a 1987 synthesis of two national 1986 studies in the United States found that race was the most significant factor in locating toxic landfills and that 3 of 5 Black and Hispanic Americans, and approximately half of all Asians, Pacific Islanders, and American Indians, lived in communities with uncontrolled toxic waste sites (Gilio-Whitaker, 2019). And while the roots of racism, misogyny, and extraction are commonly and firmly planted in dominion, actions on these matters are often pitted against one another using class, religion, nationality, and other constructs, and all and each constrained by what is deemed possible within capitalism and not overly deleterious to economic growth.

Accountability

Accountability is cited as one of the main stems of evaluation (Alkin, 2004); from the perspective of sustainability, accountability can be described as a highly evolved contagion. It is a management construct designed to enable monitoring and improvement of agreed outcomes and is usually linked to program, management, and personnel performance. Managers and programs seek to constrain risk of falling short on accountability metrics by focusing on what they have the authority, resources, and capacities to be able to likely achieve. This provides incentives to narrow the programmatic box for which they are accountable and to resist being accountable for contributing usefully to other boxes.

The two stocktakings observe that the natural system, sustainability, and the nexus are systematically absent from evaluations. However, the remit of some agencies does address the natural

system, such as the Global Environment Facility (GEF), UN Environment, national government departments such as environment and natural resources, and environmental NGOs. The evaluation record of these is mixed; while the GEF Independent Evaluation Office addresses both systems and incorporates nexus, other natural system agencies focus almost exclusively on the natural system. We can generalize this by framing evaluations as single system (either human or natural) or two system (Rowe, 2012).

Evaluations are overwhelmingly single system, a situation to which accountability frames contribute. Since natural system values are infrequently considered, accountability reinforces ignoring the natural system. We know that even within the human system we must recognize and incorporate connectivity to reach to public policy goals. Reinforcing and incentivizing partitions between human and natural systems and within human systems accountability reinforces silos, the opposite of the silo busting required for evaluation at the nexus and for evaluation more generally.

Evaluating sustainability requires evaluation practices and methods that (a) recognize and operate at the nexus where both human and natural systems are present and (b) address the intrinsic coupling between and within human and natural systems. It is bad enough that the natural system is not valued and that systems approaches and understanding are unlikely with political and administrative partitioning. Accountability reinforces and further constrains possibilities of addressing sustainability in programming and evaluation with its focus on “accountability scales” that rarely reach beyond the accountability frame of the intervention.

One result is that the responsibility and remit of the intervention and reach of its direct effects frame the spatial scale for the evaluation within the larger framing of governance structures such as local area, province, or country, or within the remits of the responsible government organization. Ecosystems and landscapes provide more relevant spatial framing for natural systems; there is no reason to expect the boundaries and shapes of ecosystems to align with human system politi-

cal and administrative boundaries or program areas. And ecosystems are not always appropriate for the territory of an organism—for example, a wolf, whale, or snail function across or entirely within an ecosystem. At a minimum, the relevant spatial scales for the natural system can be thought of as an ecosystem, often highly coupled with other ecosystems. This presumption that boundaries and territories in the natural system will align with the political and administrative boundaries of the human system is not limited to the natural system. For example, the same assumption is made about boundaries of Indigenous traditional lands, and that the Canada/United States border is relevant or appropriate where it crosses traditional lands. For many Indigenous peoples, the relevant spatial boundaries are their traditional territories from which food; medicine; and spiritual, ceremonial, and community values are drawn (Gilio-Whitaker, 2019). Instead, evaluation is likely to address the spatial scales defined by colonial occupation such as a reserve or First Nation territory; these are always and importantly smaller than traditional territories and often exclude areas of high importance to Indigenous peoples.

Program managers, evaluators, and especially evaluation commissioners often insist that an evaluation be conducted within the frame of the stated goals and operations of (accountability of) the intervention. This severs interventions from each other and limits the reach of evaluation, falling well short of the critically important public policy goals such as ending poverty or achieving sustainability. As Williams (2019) observed, such a frame establishes a program and evaluation ecosystem where programs systematically are assessed as providing positive contributions to the broad goal and where no progress is visible toward achieving the goal itself. It also creates a systematic positive bias in evaluation (Rowe, 2019b).

Institutional Capture

Institutional capture is the process by which identified needs and demands for major structural

change are captured by existing structures, policies, and approaches. The SDGs were such a moment when sustainability was recognized as an overriding priority requiring major structural change to address. By and large, responsibilities for individual SDGs were assigned without changing the partitioned structure of organizations. But successfully addressing sustainability programmatically or in evaluation requires platforms suited to the task; the partitioned structures, policies, and approaches are not well suited to pursuit and evaluation of sustainability. Understandably, the UN and other multilateral organizations staked claims on specific SDGs, pursuing the assurance this provided to their futures; some such as the United Nations Development Programme (UNDP), the United Nations Industrial Development Organization (UNIDO), and the International Fund for Agricultural Development (IFAD) now explicitly recognize this connectivity, while others are on the pathway to do so.

The evaluation criteria of the Development Assistance Committee of the Organisation for Economic Co-operation and Development (OECD DAC) Network on Development Evaluation (2019) address sustainability as sustaining interventions and achievement of impacts, and not, as most think of sustainability, as a nexus concept of human and natural systems together with emphasis on sustaining the capacity of the natural system to enable life. In this, evaluation is somewhat distinct—elsewhere sustainability is recognized as a science “with a room of its own” (Clark, 2007); the 2009 Nobel Prize in Economics was awarded to Elinor Ostrom for her work on the commons and as one of the founders of coupled human and natural systems (CHANS) analysis. And, as the UNEG and CES stocktaking efforts have shown, evaluation has also been largely captured by the institutions it serves.

Sustainability-Ready Evaluation

Evaluators are good observers and place confidence in good evidence. They will increasingly be persuaded by the emerging knowledge on sus-

tainability and climate, and increasingly recognize how these have affected the human issues and populations that have been evaluators' primary concern. They will also recognize how their long-preferred interventions and methods in the human system can contribute to worsening climate and sustainability. The underlying premise of sustainability-ready evaluation is that evaluators will recognize the need to address effects in the natural as well as the human system and take evaluation to a place where existing capacities are insufficient. Evaluators will need to, for example: recognize, speak, and hear representatives of natural system knowledge; learn how to feasibly address dynamically coupled systems (Liu, 2007); incorporate effects that have widely differing temporal and spatial scales and very differently framed units of account; and be open to and advocate for shared evaluation functions (see Carugi & Bryant, 2019; Rowe, 2012; Uitto, 2019).

Other fields of inquiry and assessment will be important contributors to developing and implementing evaluation at nexus settings. Evaluation is a cross-disciplinary field accustomed to drawing from other fields of inquiry, and this is fortunate because evaluating sustainability will require knowledge from and engagement from more system sciences. Climate and materials sciences, ecology, and geography will be important as will knowledge from more focused fields such as energy engineering, biology, agriculture, forestry and fisheries, and areas of public administration such as procurement. Two connected fields concerned with understanding and assessing nexus will likely be critically valuable fellow travelers: sustainability science (Kates, 2011; Clark et al., 2016) and CHANS work and networks (Liu, 2007; Ostrom, 1990).

Strongly siloed culture, structures, and practices of evaluation and programs create challenges to mainstreaming sustainability in the nexus sense of human and natural systems. To truly incorporate the natural system into long-standing and newer interventions whose primary

focus is in the human system is proving difficult; likewise, to get evaluations to address the natural system is challenging, as shown by the UNEG stocktaking. However, the effort does appear to be gaining some momentum, such as in research on environmental effects of refugee camps (Braun et al., 2016), although discarded Covid-19 face masks are already finding their way to landfills and water bodies. As Fabien Cousteau (2020) wrote recently,

We live in a closed-loop system. We can't actually throw things "away." The plastic we toss in the garbage often just ends up inside the bodies of marine animals, before finding its way back inside of us. (para. 12)

This means that what are usually classed as unexpected or unintended effects, or effects that were known but ignored because they lay outside the accountability frame of the intervention, now have to be recognized as a direct effect of the intervention. I have shown (Rowe, 2018, 2019a, b) that ignoring direct effects in the natural system imparts a systematic positive bias to evaluations. To make the point clear, evaluation conducted in silos has a systematic positive bias favorable to the intervention and, importantly, arising because of the accountability frames that are applied as discussed above.

Sustainability-ready evaluation is an evaluation function that is ready to recognize these connections and able to cross them. It is an evaluation function with individual evaluators and evaluation organizations that are enthused by contributing to a future we choose (Figueres & Rivett-Carnac, 2020). There are many strongly held visions of what that future should be, with associated and strongly held views of what we need to change to get there. It is not the job of evaluation to pick a pathway or end point; our job is to be enthused and capable of contributing to improvement, including sorting and valuing the competing pathways and desired new ways. Evaluation today is appropriately described as close to sustainability-ignorant and far from sustainability-ready.

How Can Evaluation Contribute to Checkmating Extinction?

An evaluation able to contribute to the defeat of extinction requires some relatively simple changes in how we frame and undertake our work, but these simple changes will significantly alter the stance and thus the politics of evaluation. Here I briefly sketch some important changes in stance for an evaluation fit for purpose for the endgame.

Checkmating extinction will only be possible if evaluation shifts from a singular focus on the human system to mainstreaming nexus in all evaluation. We are right now at a juncture where urgently needed changes seem possible. A second major change in the stance of evaluation relates to expectations of goal achievement: The current standard of progressing toward goals merely draws out a checkmate in favor of extinction. Instead, overcoming extinction requires a stance at the endpoint and assesses achievement of these goals with evaluation providing guidance to improve performance. Of course, achieving these goals requires joined-up, system-wide efforts for which we need to join evaluation stances with systems approaches. Conditions are worsening faster than expected and efforts to understand status and trends in natural systems and options for mitigation and adaptation are generating new knowledge at a rapid pace. This means that the stance of evaluation must be nimble and adaptive to integrate these changes, and be undertaken with sufficient rapidity to align with significantly accelerated decision cycles. Together, all of this means that evaluation for the endgame must be relentlessly use seeking and forward looking.

These are but some of the features needed for an evaluation function and practice that is an ally in efforts to checkmate extinction. Consideration of this stance will identify additional necessary features and perhaps diminish the importance of some that are discussed below. This chapter is only an early step in identifying the stance needed for an evaluation that contributes to the endgame.

Recognizing Natural Systems as the Foundation for the Human System Means Adding the Natural System Perspective to All Evaluation Criteria

The opponent at the endgame is continued destruction of the natural system by humans, meaning that both systems must be considered and addressed by evaluation at the endgame. That is, nexus is the required position for evaluation at the endgame.

Think of the relationship between human and natural systems with the natural system as a bank account. The human system has well exceeded its overdraft limit so that now every draw we make must have a repayment schedule that not only matches current withdrawals but also systematically and strategically starts to reduce the overdraft.

Environmental and social safeguards and policies have been enacted by most development donors with the requirement that they are applied in project development, funding, implementation, operation, and assessment (IFAD, 2018; World Bank, 2020). These standards are relatively recent, most enacted in the past decade, and the documents clearly consider human and natural systems as connected. In practice, however, climate and environment/natural resource management are usually treated as additional criteria that must be addressed in project design and assessment, isolated and marginalized rather than imbedded into planning.

We can consider inclusion of the natural system criteria in four phases, defined by requirements to meet the threshold to achieve a “satisfactory” rating:

1. Ignored: In this phase, environment (and climate) were rarely addressed, development was the priority and equity issues were important. **Result: Increase in the overdraft on the natural system account.**
2. Good intentions: Environment and climate were noted in this second phase, often with what could be described as a faith-based

approach. It was not unusual to see project designs, evaluations, and supervision reports that considered commitments to compliance with donor environmental guidelines and safeguards and with national regulations to warrant a satisfactory rating. To put this in perspective, I have never seen an evaluation of an education or health intervention make a statement such as, “The design of the intervention incorporated government guidelines and a designated body has the authority to inspect and enforce, so we deem the approach satisfactory.” Substitute environment for education in the previous sentence and we have a statement that is frequently made about natural resources, sustainability, and climate in supervision and evaluation documents. **Result: Increase in the overdraft on the natural system account.**

3. Do no harm: With these emerging approaches, achieving a satisfactory rating for climate and environment requires plausible design and implementation resources and responsibilities such that the intervention will not harm the environment or ignore climate. Empirical evidence is not required for a satisfactory rating but might become an expectation. Use of less harmful practices for continued resource extraction, such as climate-smart agriculture, species-specific fishing gear, protection of mangroves, forest management, and methods in road building, are deemed to not harm and so warrant a satisfactory rating. In effect, this is a type of double counting with the natural system benefits, such as improved irrigation and soil condition, required to restore production levels and support previous harmful agricultural projection practices. **Result: End of continued withdrawals on the natural system account but accumulated overdraft not addressed.**
4. Evaluation we need: In the fourth phase, evaluation for the endgame, achieving a satisfactory rating requires that restorative actions for the natural system are confirmable, central, and substantial parts of project design, operations, and adaptive management. **Result: Paying down the overdraft; learning and dif-**

fusion provide positive prospects that this will continue and accelerate.

Mainstreaming sustainability systematically locates evaluation at the nexus and is a first and essential change in the stance of evaluation; but valuing the natural system evaluation is beginning to address dominion.

Evaluation Standards Will Emphasize Achieving the Larger Goals Identified as Central to Checkmating Extinction

When the end is in sight, when the endgame is what is at play, our focus shifts from playing the game well (admirable evaluation) and from contributing to incremental improvements for beneficiaries to an absolute need to provide value to checkmating our destruction of the natural system that sustains us.

To illustrate the character of absolute evaluation standards, the International Resources Panel (IRP)⁷ has shown that the planet does not have the material resources to provide for expansion of existing cities and creation of new ones resulting from urbanization, rural-to-urban migration, and population increase (Swilling, 2018). Development projects typically claim they will “contribute to” slowing rural-to-urban migration through improved rural livelihoods. Rural-to-urban migration and population growth are complicated and involve a powerful mix of push and pull factors requiring combined programmatic efforts to achieve sustainable flows and levels that will contribute to sustainable development, not undermine it. This is an illustration of a goal important for the endgame; evaluation needs to assess against achievement of that goal. If population and urban growth threaten sustainability, then the standard that needs to be applied in evaluation is achieving the goals that will remove population increase and rural-to-urban migration as important threats to sustainability. This does not mean curtailing migration and mobility, which are important to escaping severe climate

⁷<https://www.resourcepanel.org/>

and for humanitarian and economic reasons. Achieving levels of rural-to-urban migration sustainable for both urban and rural areas likely hinges on viable rural communities. And evaluation can provide value in moving from current unsustainable flows by adopting a stance that includes an expectation of verifiable achievement of important endgame outcomes that will realize specific migration goals set at sustainable levels. These goals, like the high-end climate goals of a CO₂ reduction to limit temperature increase to 1.5°C, should be specified in absolute terms; for example, the specific sustainable population of Vancouver or Hanoi.

Standards Need to Shift to Evaluating Against Collective Achievement of Sustainability Goals, and Away from Likely Contributions by Partitioned Organizations and Interventions

Achieving the results needed to checkmate extinction requires collective and synthesized efforts; this is the required stance of evaluation for the endgame.

Partitions must be replaced by joined-up action and evaluation must adopt a collaborative focus on system achievement of the larger goals required for sustainability, regardless of whether interventions have adopted this stance. Holding interventions accountable for achieving results for which they are neither resourced nor authorized is inappropriate. However, for the endgame, evaluators should still address the needed result, what is required to achieve it, and the success and contributions of efforts toward collectively addressing this result. Setting goals that are critically important to success in the endgame is one way evaluation can observe shortcomings in collaboration and shared efforts toward achievement. It will also reveal gaps between current and needed achievements that likely span a number of individual organizational remits. This type of evaluation, focusing on what is needed, reflects the spirit of a results focus but from a collective, joined-up perspective rather than from parti-

tioned efforts. It promotes collective action and accountability for sustainability goals.

Sustainability Is Imbedded in All Evaluation Criteria Reflecting Nexus, Not Isolated as a Free-Standing Criterion

An evaluation stance recognizing the complex connectivity of human and natural systems means that all evaluation criteria should be considered from a two-system stance—sustainability and climate should not be isolated in separate and usually marginalized criteria.

Collective action means that work toward any and all of the SDGs and government and third-sector initiatives is likely to be drawing from and contributing to the sustainability of the natural system and climate. The previous element brings these into the scope of evaluation for the endgame, and this element addresses how evaluation accomplishes this. Each of the evaluation criteria and standards, e.g. the OECD DAC criteria (relevance, coherence, effectiveness, efficiency, impact, and sustainability), needs to be infused with considerations of sustainability by addressing both human and natural systems. Examples include the effect of humanitarian efforts on the physical landscape, and the many effects on the natural system of the use of plastics.

Evaluation Standards at the Endgame: Evaluating with Rapid Change and Uncertainty

Relentless rapid learning and brisk adaptation is the temporal scale required for interventions at the endgame and so must also be for evaluation.

Sustainability and climate are topics where the knowledge and practice base is improving rapidly and still features considerable uncertainty and ambiguity. Where changes in our knowledge are proceeding at a rapid pace and where considerable ambiguity still exists, longer term interventions—such as 4 or more years—will inevitably be suboptimal by the time they are

halfway through their remit, perhaps highly sub-optimal. Those implementing interventions must adopt vigorous adaptive management practices and be held accountable for this. We need to accelerate the pace of reflection and renewal, or else an important portion of our efforts will be applying approaches that are no longer considered efficacious at a time when we can least afford to do so. Evaluation is an important vehicle for this.

At the endgame, knowledge cycles are greatly reduced—we now think that the shelf life of some current climate knowledge is about 2 years. Severe climate events are also accelerating, becoming more frequent and severe and building cumulative effects. Two category 5 storms and resulting flooding within one month, as happened in 2020 in the Caribbean, requires very different responses than two storms of equal strength separated by 10 years.

To illustrate, consider 2030 and 2050 as forecasts of when we will pass irreversible thresholds, which make them key timings for checkmating extinction. A large portion of program and project cycles approach 7 or more years from inception to renewal, with 1–2 years for planning and funding, 1–2 years to mobilize, then operations of 4–5 years. Seven-year program cycles gives us just one program cycle until 2030 and four until 2050. The typical mid-term, end-of-term, and later ex-post evaluation approaches cannot provide information, insights, and advice in time to affect interventions in much more rapid adaptation cycles. Some evaluation approaches and methods will need to adapt rapidly and significantly to be relevant to evaluation at the endgame; fortunately, other approaches and methods are more fit for this purpose. Longer term evaluation undertakings will still provide value, such as with longer term impacts and adaptation of interventions to changing conditions, but overall, evaluation at the endgame is a new challenge for the field, requiring the evaluation stance to immediately become shorter term and employ more rapid approaches that are

relentlessly use seeking such as Rapid Impact Evaluation (Rowe, 2019a).

Evaluation for the Endgame Relentlessly Pursues Use

We no longer have the luxury to indulge the evaluation agendas and strategies that do not contribute to checkmating extinction. Our work must focus directly and strongly on the rapid adaptation and learning cycles of a proliferating landscape of actions contributing (or not) to checkmating extinction.

Conclusion: Nexus Requires New Rootstock to Grow Relevant Evaluation Functions

This chapter recognizes that we have entered the endgame of extinction and identifies what is needed for evaluation to contribute to checkmating extinction. I have sketched a trail from where evaluation is today to where it needs to be to provide value and guidance to efforts to achieve a checkmate favorable to life on the planet.

That trail first observes that evaluation at global and national levels is monastically focused on the human system and only marginally addresses the natural system. It reaches back to Judeo-Christian concepts of dominion as the origin story for our focus, and identifies narrowly framed accountability structures as an important contemporary mechanism for the exercise of dominion. Reinforcing this is institutional capture of efforts to infuse sustainability and systematically address necessary climate goals in development and associated social ambitions at all levels. The unhappy result is seen in two recent stocktaking efforts illustrating the limited contributions of contemporary evaluation to sustainability.

Evaluation at the endgame is different from the evaluation we have known and practiced up to

now. Evaluation will need to take stances that will be challenging, as is any endgame effort. The six characteristics of evaluation for the endgame are:

1. The opponent at the endgame is continued destruction of the natural system by humans, meaning that both systems must be considered and addressed by evaluation at the endgame. Nexus is the required position for evaluation at the endgame.
2. When the end is in sight, our focus shifts from playing the game well (admirable evaluation) and from contributing to incremental improvements for beneficiaries to an absolute need to provide value to checkmating our destruction of the natural system that sustains us.
3. Achieving the results needed to checkmate extinction requires collective and synthesized effort, which is the required stance of evaluation for the endgame.
4. An evaluation stance recognizing the complex connectivity of human and natural systems means that all evaluation criteria should be considered from a two-system stance—sustainability and climate should not be isolated in separate and usually marginalized criteria.
5. Relentless rapid learning and brisk adaptation is the temporal scale required for interventions at the endgame and so must also be for evaluation.
6. We no longer have the luxury to indulge the evaluation agendas and strategies that do not contribute to checkmating extinction and our work must focus directly and strongly on the rapid adaptation and learning cycles of a proliferating landscape of actions contributing (or not) to checkmating extinction.

Adopting these stances at first appears to be a radical shift for evaluation, one with poor prospects for adoption. However, a growing recognition of the sustainability and climate imperative is underway. Evaluation working with biophysical knowledge partners is able right now to usefully contribute to the endgame. The hard part is recognizing that the prevailing stance of evaluation is contributing to the problem, that we need

to turn our backs on forces and institutional arrangements that have provided us comfort in exchange for complicity, and turn to a future we choose, which is to be a valued and useful contributor to checkmating extinction.

References

- Adyel, T. M. (2020). Accumulation of plastic waste during COVID-19. *Science*, 369(6509), 1314–1315. <https://doi.org/10.1126/science.abd9925>.
- Alkin, M. C. (2004). *Evaluation roots: Tracing theorists' views and influences*. Sage.
- Better Evaluation. (2020). *Footprint evaluation*. https://www.betterevaluation.org/en/themes/footprint_evaluation
- Braun, A., Lang, S., & Hochschild, V. (2016). Impact of refugee camps on their environment: A Case study using multi-temporal SAR data. *Journal of Geography, Environment and Earth Science International*, 42(2), 1–17. <https://doi.org/10.9734/JGEEIS/2016/22392>.
- Carugi, C., & Bryant, H. (2019). A joint evaluation with lessons for the Sustainable Development Goals era: The joint GEF-UNDP evaluation of the Small Grants Programme. *American Journal of Evaluation*, 41(2), 182–200. <https://doi.org/10.1177/1098214019865936>.
- Chelimsky, E. (2012). Public-interest values and program sustainability: Some implications for evaluation practice. *American Journal of Evaluation*, 35(4), 527–542. <https://doi.org/10.1177/1098214014549068>.
- Christie, C. A., & Alkin, M. C. (2008). Evaluation theory tree re-examined. *Studies in Educational Evaluation*, 34(3), 131–135. <https://doi.org/10.1016/j.stueduc.2008.07.001>.
- Clark, W. C. (2007). Sustainability science: A room of its own. *PNAS*, 104(6), 1737–1738. <https://doi.org/10.1073/pnas.0611291104>.
- Clark, W. C., van Kerkhoff, L., Lebel, L., & Gallopin, G. C. (2016). *Crafting usable knowledge for sustainable development*. National Academy of Sciences.
- Cousteau, F. (2020, December 9). Our oceans, our future. *New York Times*. <https://www.nytimes.com/2020/12/09/opinion/covid-climate-change-ocean.html>
- Figueres, C., & Rivett-Carnac, T. (2020). *The future we choose: Surviving the climate crisis*. Knopf.
- Gilio-Whitaker, D. (2019). *As long as grass grows: Indigenous fight for environmental justice, from colonization to Standing Rock*. Beacon.
- International Fund for Agricultural Development. (2018). *IFAD's social, environmental and climate assessment procedures (SECAP)* (2017 ed.). Author.
- Kates, R. W. (2011). What kind of a science is sustainability science? *PNAS*, 108(49), 19449–19450. <https://doi.org/10.1073/pnas.1116097108>.

- Liu, J. D. (2007). Coupled human and natural systems. *Ambio*, 36(8), 639–649. [https://doi.org/10.1579/0044-7447\(2007\)36\[639:CHANS\]2.0.CO;2](https://doi.org/10.1579/0044-7447(2007)36[639:CHANS]2.0.CO;2).
- Nadeau, R. L. (2006). *The environmental endgame*. Rutgers.
- Nadeau, R. (2008, March 19). Brother, can you spare me a planet? (extended version). *Scientific American*. <https://www.scientificamerican.com/article/brother-can-you-spare-me-a-planet/>
- OECD/DAC Network on Development Evaluation. (2019). *Better criteria for better evaluation*. <https://www.oecd.org/dac/evaluation/daccriteriaforevaluatingdevelopmentassistance.htm>
- Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action: Political economy of institutions and decisions*. Cambridge University Press.
- Patton, M. Q. (2020a). *Blue Marble evaluation*. Guilford.
- Patton, M. Q. (2020b). Evaluation criteria for evaluating transformation: Implications for the coronavirus pandemic and the global climate emergency. *American Journal of Evaluation*, 1–37. <https://doi.org/10.1177/1098214020933689>.
- Rowe, A. (2012). Evaluation of natural resource interventions. *American Journal of Evaluation*, 33(3), 384–394. <https://doi.org/10.1177/1098214012440026>.
- Rowe, A. (2018). Ecological thinking as a route to sustainability-ready evaluation. In R. Hopson & F. Cram (Eds.), *Tackling wicked problems in complex ecologies* (pp. 25–44). Stanford University Press.
- Rowe, A. (2019a). Rapid impact evaluation. *Evaluation*, 25(4), 496–513. <https://doi.org/10.1177/1356389019870213>.
- Rowe, A. (2019b). Sustainability-ready evaluation: A call to action. *New Directions in Evaluation*, 162, 29–48. <https://doi.org/10.1002/ev.20365>.
- Sustainability Working Group, Canadian Evaluation Society. (2020). *Report on stocktaking for sustainability-ready evaluation (draft)*. (Unpublished manuscript).
- Swilling, M. H. (2018). *The weight of cities: Resource requirements of future urbanization*. United Nations Environmental Programme, International Resources Panel. <https://www.resourcepanel.org/reports/weight-cities>.
- World Bank. (2020). *The environmental and social framework*. <https://projects.worldbank.org/en/projects-operations/environmental-and-social-policies#safeguards>
- United Nations Evaluation Group Working Group on Integrating Environmental and Social Impact into Evaluations. (2020). *Stock-taking exercise on policies and guidance of UN agencies in support of evaluation of social and environmental considerations, Vol I Main Report*. United Nations Evaluation Group. <http://www.unevaluation.org/document/download/3712>.
- Uitto, J. I. (2019). Sustainable development evaluation: Understanding the nexus of natural and human systems. *New Directions in Evaluation*, 2019(162), 49–67. <https://doi.org/10.1002/ev.20364>.
- Williams, R. (2019). Evaluation in a dangerous time: Reflections on 4 years in a central policy agency in the Government of Nova Scotia. *Evaluation Matters—He Take Tō Te Aromatawai*, 5, 41–62. <https://doi.org/10.18296/em.0039>.
- Zak, D. (2019). How should we talk about what’s happening to our planet? *Washington Post*, August 27. https://www.washingtonpost.com/lifestyle/style/how-should-we-talk-about-whats-happening-to-our-planet/2019/08/26/d28c4bcc-b213-11e9-8f6c-7828e68cb15f_story.html

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

