Article

Evaluation
1–25

DOI: 10.1177/13563890251363232
journals.sagepub.com/home/evi
Sage

A critical, integrative review on evaluating systems change and transformation, Part One: 2011–2021

Emily F. Gates
Boston College, USA

Pablo Videira
Comillas Pontifical University, Spain

Maya Komakhidze
Boston College, USA

Charlotte Aldrich
Boston College, USA

Clara Shim
Boston College, USA

Abstract

Calls for evaluating systems change and transformation in multiple fields present an opportunity to explore cross-field patterns. This article reports on part one of a critical, integrative review of academic and gray literature published between 2011 and 2021 (n = 102) within five areas: evaluation, health, organizational change, sustainability, and philanthropy. Questions address key definitions, how a systemic approach differs from traditional social problem-solving, leverage points to influence change, and implications for evaluation. Four findings include (1) limited normative debate about change and transformation; (2) conventional and systemic approaches contrasted as binary paths; (3) 10 shared leverage areas with the least attention on power and resources; and (4) an expanded role for evaluation that presents challenges and opportunities. Review results provide support for shifts underway within the evaluation field, including funders

Corresponding author:
Emily F. Gates, Boston College, 140 Commonwealth Avenue, Chestnut Hill, MA 02467, USA.
Email: emily.gates@bc.edu

working in deeper collaborations, evaluators expanding their skill sets, and intermediary agencies facilitating transdisciplinary exchanges.

Keywords
evaluation, large-scale change, learning, monitoring, systems change, transformation

Introduction

Given global calls to transform evaluation to support systems change and transformation, scholars Ofir and Rugg (2021: 47) describe a critical juncture:

> Evaluation specialists today carry the responsibility to revisit, redesign, and reconfigure, with a sense of urgency, evaluation theories, and practices to be valuable and essential instruments in support of the large-scale, transformative changes our societies and ecosystems need. Unless we can show that evaluation can powerfully contribute to the new world humanity has to create, it is better that our craft slowly fades into irrelevance rather than being a persistent obstruction to the transformations needed.

Numerous evaluation scholars and international organizations echo this sentiment (Naidoo, 2022; Patton, 2021; Picciotto, 2020; Van den Berg et al., 2019). Transforming evaluation challenges the field's historical and long-standing focus on assessing the effectiveness of discrete projects, policies, and change initiatives using linear and mechanistic logic models and designs premised on counterfactual and attribution-based causation (Schwandt and Gates, 2021). Instead, evaluation increasingly needs to incorporate systems thinking and complexity science and methods (Gates et al., 2021; Barbrook-Johnson et al., 2021; Patton, 2019) and forward equity, social justice, and decolonization (Dean-Coffey and Coné, 2023; Mertens, 2010).

Critiques of evaluation and calls for its transformation in multiple fields are an opportunity to explore cross-field patterns (Patton, 2023). Despite buzz around the terms "systems change" and "transformation," there has been little empirical research on the topic within the evaluation field, to date, and a wide range of work, often using distinct terminology, in related fields. While some insights come from field-specific reviews in public health, medicine, and nursing (e.g. Best et al., 2012; Braithwaite et al., 2018; Cristancho et al., 2018; Olsson et al., 2020; Rusoja et al., 2018), there is an opportunity to explore cross-field patterns, differences, and gaps.

This article reports part one of a critical, integrative review of academic and gray literature on systems change and transformation published from 2011 to 2021 (n = 102) within five areas: evaluation, health, organizational change, sustainability, and philanthropy. Materials reviewed included 87 peer-reviewed journal articles and 15 non-academic, practice-focused references. Research questions address key definitions; how a systemic approach differs from conventional social problem-solving; leverage points to influence change; and implications for evaluation. Our analytic process consisted of two rounds of coding and interpretive analysis by the entire research team, conceptual and narrative synthesis led by the lead author with team consultation, and critical appraisal by the first two authors.

Results include four overarching findings: (1) minimal normative debate about change and transformation; (2) conventional and systemic approaches contrasted as binary paths rather than a middle way; (3) 10 shared leverage areas with least attention on power and resources;

and (4) an expanded role for evaluation that presents challenges and opportunities. For each finding, we provide our critical appraisal followed by integrative summaries of the reviewed literature. We then situate these findings within ongoing and recent evaluation debates and consider their implications for shifts underway in the evaluation field among funders working collaboratively, evaluators expanding their skillsets, and intermediary agencies facilitating transdisciplinary exchanges.

Methods

We conducted part one of a critical, integrative review of literature on systems change and transformation in five areas: evaluation, health, organizational change, sustainability, and philanthropy. Important to note, this is not a review of the evaluation literature; rather evaluation is one of five fields included. Our methods draw from domain-based reviews that seek to synthesize and extend a body of literature within a substantive area (Palmatier et al., 2017) and state-of-the-art reviews that address current matters to offer new perspectives and areas for further research (Grant and Booth, 2009). Conceptual and narrative integration was our goal, as we sought to identify cross-field patterns and critically appraise results to inform future directions (Grant and Booth, 2009).

Research questions

Four questions guided this review: (1) How do reference authors define and use the terms "systems change" and "transformation"? (2) How does systems change and transformation differ from traditional approaches? (3) What are key leverage points to influence systems change or transformation? (4) What are the implications for monitoring, evaluating, and learning?

Author expertise and motivation

Team members brought international and multidisciplinary perspectives to this review. Our vantage points are shaped by our work in the United States, Spain, Georgia, and Korea, with disciplinary training and/or professional experience in education, public health, engineering, and environmental studies. The two lead team members have extensive knowledge of evaluation scholarship, particularly in US and European contexts, and the systems and complexity sciences. Together, we embodied a "critical friend" role toward the literature, balancing opportunism with skepticism. For personal reasons, we were delayed in finalizing this review for publication. To address this, we position this review as part one and are currently completing part two which includes references from 2022 to 2025. Where possible, we situate review findings in recent evaluation scholarship and practice trends.

Search strategy

We sought to include significant references within the emerging area of evaluating systems change and transformation, including academic, peer-reviewed articles and gray literature (i.e. produced and/or disseminated outside scholarly, peer-reviewed publications). We used a combination of Google Scholar, targeted journal searches, Google, and researcher-added based on

Table 1. Targeted journals searched.

Area	Journal names
Evaluation	*Evaluation*
	American Journal of Evaluation
	Evaluation and Program Planning—Research Evaluation
	Performance Evaluation
	Journal of Multidisciplinary Evaluation
	Educational Evaluation and Policy Analysis
	Evaluation Review
	New Directions for Evaluation
	Canadian Journal of Program Evaluation
	Evaluation Journal of Australasia
	Practical Assessment, Research, and Evaluation
	Evaluation and the Health Professions
	Evaluation & Research in Education
	Measurement and Evaluation in Counseling and Development
	Educational Research and Evaluation
	Studies in Educational Evaluation
	Searched but excluded:
	African Evaluation Journal (0 relevant hits)
	Japanese Journal of Evaluation Studies (0 relevant hits)
	LeGes Swiss Evaluation Society—German/French, excluded
Systems & Complexity	*Emergence: Complexity and Organization*
	International Society for the Systems Sciences Proceedings
	Systems Research and Behavioral Science
Change	*Journal of Change Management*

a review of reference lists in included studies. We used the primary search terms [system* OR complex*] AND [change OR transformation OR large-scale] AND [evaluat*] within the title or abstract. We conducted searches in 2021 and early 2022 limited to works published in English between 2011 and 2021. This choice of dates built on prior reviews (Gates, 2016) and encompassed a decade of substantive development. Google Scholar searches yielded peer-reviewed academic articles, book chapters, and books. Google search expanded our search to non-academic and practice-focused pieces (e.g. guides, reports, white papers), generating a total of 240 relevant works. Finally, we searched specific journals to ensure that we had papers in focal areas of scholarship and identified additional relevant references from reference lists during our review. Table 1 lists journal searched.

Eligibility criteria and selection

We included references for which the title and/or abstract discussed theory, methodology, practical guidance, and/or case studies conceptualizing, designing, leading, implementing, and/or evaluating initiatives to change and/or transform systems or large-scale systems. During initial searches and screening of titles and abstracts, we excluded books, blog posts, social media posts, and systems references without a theoretical or methodological basis (e.g., education system, healthcare system). We also left out case studies without broader discussion of theoretical, conceptual, or methodological issues and implications. After reviewing titles and abstracts, we further excluded references that did not directly discuss systems change or transformation. See Figure 1 for search and selection process.

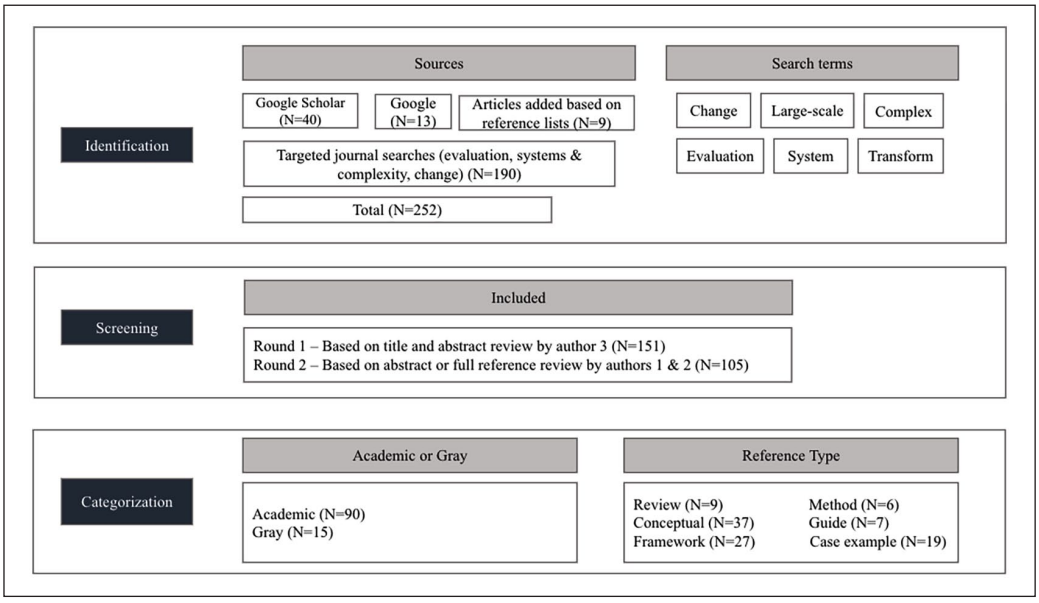

Figure 1. Reference search and selection flow diagram.

Table 2. Descriptive categorization of included references by domain.

Field	%	Academic or Gray		Reference Type					
		Academic	Gray	Review	Conceptual	Framework	Method	Guide	Case Example
Evaluation (n = 30)	29%	28	2	2	10	11	4	1	2
Health (n = 27)	26%	27	0	2	3	7	1	1	13
Organizational change (n = 9)	9%	8	1	0	6	2	1	0	0
Philanthropy (n = 12)	12%	1	11	1	5	1	0	5	0
Sustainability (n = 24)	24%	23	1	9	4	6	0	0	5
Total (n = 102)		87	15	14	28	27	6	7	20
	100%	85%	15%	14%	27%	26%	6%	7%	20%

Descriptive categorization and analysis

We developed a literature database containing key information extracted from the references and additional descriptors added by our team. A total of 102 references: 85% corresponded to academic literature, while 15% to gray literature. In terms of fields, Evaluation, Health, and Sustainability encompass almost 80% of the references reviewed (29%, 26%, and 24%, respectively). Regarding reference types, conceptual papers and proposed frameworks comprise more than half of the references reviewed (27% and 26%, respectively), followed by case studies (20%). Review papers, methodological proposals, and guides encompassed the remaining 27% of the documents reviewed. Table 2 describes references included by domain.

We employed a multi-step process to analyze each reference and synthesize the literature to answer the research questions (Whittemore and Knafl, 2005). We divided the selected literature by field and team member. We then developed, tested, and revised a codebook, which we applied using MaxQDA, a qualitative data analysis software (https://www.maxqda.com/). Deductive codes came from our research questions, such as definitions of key terms, levers for change, and the role of monitoring, evaluation, and learning (MEL). We added inductive codes throughout the process if multiple team members saw the relevance of the code based on the field(s) they were reviewing. We read and coded each reference using this shared codebook within MaxQDA in separate files, which we merged.

With coding completed, each team member developed analytic memos to answer research questions about the field(s) they reviewed (i.e. reading coded data excerpts and drafting summary sentences and bullet points in a shared analytic document). Using this document, the team reduced 102 references to a 125-page summary document upon which to base our synthesis.

Critical, integrative synthesis

The first two authors led the process to integrate and critically appraise results by research question. To integrate results, we used a combination of conceptual and narrative synthesis supported by tables and figures. To critically appraise results, we drew on critical lenses inspired by a phronetic approach and critical systems thinking. A phronetic approach asks: Where are we going? Who gains and who loses, and by which mechanisms of power? Is this development desirable? What, if anything, should we do about it? (Flyvbjerg, 2001). Critical systems thinking interrogates boundaries and their implications for who or what is included, excluded, and marginalized, especially in terms of potential risks and harm (Gates, 2018; Jackson, 2019; Ulrich, 1983).

Study limitations

This study has several limitations due to the time frame, included fields and reference materials, and analytic process. First, included references were published during 2011-2021. The review does not give a historical account of developments during this time period as we did not systematically search by year but instead within this timeframe. Recent references published in the last several years, from 2022 to 2025, are excluded. As a team, we conducted the review in 2022, but took time to complete and submit this written article for publication posing a substantial limitation. Second, the review includes five fields and English-language academic and gray references. Search terms centered on the topic, evaluation, systems change, and transformation and were not field-specific (e.g. health, philanthropy). This means included references do not fully represent the literature on this topic published in each field during this timeframe; instead, they provide a partial snapshot. This also means relevant literature in the evaluation field was excluded if it did not meet the inclusion criteria limiting our coverage of historical and current debates related to these topics. Our search and selection process does not allow comparisons between fields or specific claims about a field. The combined search terms and timeframe exclude fields with extensive literature published before 2011, such as the natural sciences. Omitting books further bounds findings to the reviewed references and not reflective of the full bodies of knowledge on this topic in each field. Third, our analytic process did

not evenly weigh references; references that provided more information on a research question were coded more often and, in turn, inform the findings more than those with less coded material. In generating overarching findings, we took a critical approach, leveraging our perspectives to interpret the results; other researchers may have interpreted the results differently or amplified other patterns (Grant and Booth, 2009).

Critical integrative synthesis

Here we present four overarching findings framed by our research questions. For each, we begin with our critical appraisal followed by a summary of integrative results using narratives, figures, and/or tables. To maintain readability within a limited word count, we provide key points and illustrative quotations, not the comprehensive results.

Limited normative discussion about systems change and transformation

Change, systems change, and transformation are central to the reviewed references. Yet, most authors did not directly define these terms. Instead, systems change and transformation were framed as universally desirable processes that will bring inarguable benefits, with little discussion of the underpinning values and normative questions (for an exception that takes a sociological perspective, see Jones et al., 2019). Systems change processes alter those who benefit and are harmed by how systems operate and are not "inherently good or bad" (Abercrombie, 2015: 11). Change, as deviation from a current path, function, or structure, raises normative questions about who should and should not benefit and how. Transformation further means fundamental, irreversible system altering, raising the stakes and potential consequences. Presumption of universality ignores the normative and political debates inherent in change and transformation, and practical limitations, as those who have been benefiting may oppose, resist, or redirect change efforts.

This section delineates the concepts of and purposes for change, systems change, and transformation, drawing on the literature reviewed. Introductions of papers often state one or several purposes briefly, as an inherently good or positive thing. Across the references, we identified six purposes within opening sections: sustainability, self-organization, adaptation, resilience, transformation, and equity/justice. Often, authors use multiple terms together to convey broad desires without specific meanings. For example, one reference discussed how greater *resilience* emerges from mastering challenges and reinforces shared learning, *adaptation*, and *self-organization* (Sturmberg et al., 2012: 4), stating elsewhere that "Everyone wants a *sustainable* well-functioning health system" (Sturmberg et al., 2012: 1). As illustrated here, these terms are used together in a normative way to refer to desired ends societies should strive for as they work to change or transform systems. Next, we synthesize different meanings of each term, also summarized in Table 3.

Change, an intentional deviation from the underlying path (Burch et al., 2014), is a process characterized differently by authors. Nyström et al. (2013) describe change processes from predictable to complex, uniform to disseminated and differentiated, linear to rational, and planned or episodic to emergent or continuous. Other authors propose more narrowly defined characteristics such as messy, non-linear, and often unpredictable (Narayanan and Adams, 2016), and happening at multiple scales (Burch et al., 2014). A distinction between incremental and transformative change is made across fields while some authors acknowledge that incremental change can have profound implications for transformative processes.

Table 3. Key terms and meanings.

Term	Meanings
Change	Intentional deviation from an underlying path, often toward improved outcomes or altered states. Across fields, change manifests in multiple forms with varying temporalities, degrees, and depths, including incremental adjustments, reform, or more radical transformation.
Systems change	A shift in the multiple, interconnected conditions, structures, and dynamics that sustain current patterns or problems. It encompasses change within and across systems, often through new relationships, narratives, and practices among actors and sectors. In evaluation and related fields, it is increasingly associated with an orientation that combines systems awareness, equity and justice considerations, and transformative ambition.
Transformation	A fundamental reorganization of systems resulting in new forms and functions, often triggered by critiques of current conditions and power structures. It involves emergent outcomes and contested values, and is frequently framed as necessary for responding to entrenched, complex, or crisis-laden challenges. While often treated as desirable, transformation can generate both positive and negative consequences, depending on the perspectives and positions of those affected.

Systems change is used to define a systems-wide change process or endeavor, what is subject to be changed (Burch et al., 2014), and the type of change: "(continual) emergence of new patterns of organizing" (Birney, 2021: 752). Systems change is also presented as "a combination of transformative ambition and systems awareness" (Australian Centre for Social Innovation, 2019: 7), highlighting a contrast to how change is typically pursued (Putnam-Walkerly, 2017). Cabaj (2018) and Scally et al. (2020), drawing on the work by (Kania, Kramer, and Senge, 2018: 3), focus on how systems change means "shifting the multiple conditions that hold problems in place" through changing policies and practices, resource flows, relationship, power dynamics, and mental models. Loorbach et al. (2017: 7) unpack systems change as "the result of an interplay of a variety of changes at different levels and in different domains that somehow interact and reinforce each other to produce a fundamental qualitative change in a societal system." Abercrombie (2015: 11) highlights systems change as "an intentional process designed to alter the status quo . . . [that] require[s] a radical change in people's attitude as well as in the ways people work."

Transformation involves a profound reorganization of a system's structures, properties, and controls (Andrachuk and Armitage, 2015) that leads to a new form or function of the system, and consequently to a fundamentally new system (Barnes et al., 2017) that is qualitatively different (Andrachuk and Armitage, 2015). Transformation is triggered by the inability of the current system to address entrenched challenges within its current state or regime, and thus its untenability (Barnes et al., 2017; Lee and Waddock, 2021). The concept of transformation offers compelling language for interdisciplinary collaboration and presents opportunities to address power imbalances and sources of vulnerability (Andrachuk and Armitage, 2015). Furthermore, cognitive transformation or mind-set change is an integral part of the process, either coupled with (Andrachuk and Armitage, 2015; Van Bruggen et al., 2019) or distinct from but linked to the broader transformation process (Lee and Waddock, 2021).

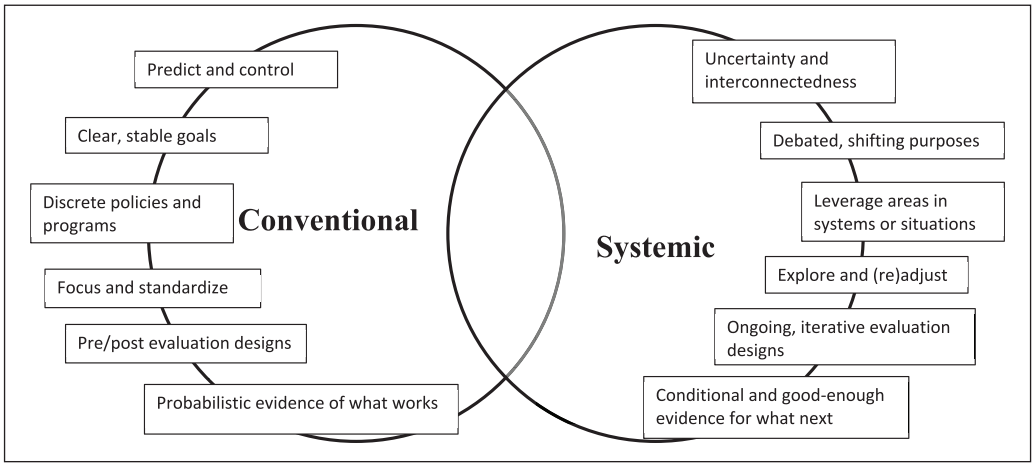

Figure 2. Venn diagram contrasting linear and adaptive assumptions.

Conventional versus systemic approach: Binary paths or a middle way?

Across the fields, a shared narrative contrasts conventional and systemic approaches to change. Authors position systems change and transformation as an alternative to traditional ways of designing, implementing, and evaluating interventions. Examples of contrasting approaches include Newtonian science to complexity and systems science, command-and-control management to adaptive management, program and policy evaluation to systems, and complexity-informed evaluation. When making this contrast, most literature emphasized these as binary paths, each fundamentally unique. A conventional approach, also called incremental, emphasizes problem-solving through piecemeal interventions that focus on a single symptom or pattern of the system regardless of how connected it is to others and irrespective of fundamental causes. A systemic approach highlights change as continuous amid complexity; intervening requires systemic awareness, codesign, and ongoing learning. Systems change and transformation are often framed either as fundamentally at odds with linear, incremental approaches (Van Bruggen et al., 2019) or at the opposite end of a continuum (Palomo et al., 2021).

However, combining conventional and systemic approaches may be necessary. For example, systems change and transformation may serve as loose framing for initiatives to optimize connections, coherence, and coordination to amplify intended positive impacts or genuinely bring about the desired system change (Waddock and Waddell, 2021). Open questions challenge the idea of binary paths, with some actors exploring new, hybrid/middle ways that beneficially blends linear/incremental and adaptive approaches. Targeted initiatives could pave the way for transformation to emerge (Westley et al., 2011). There is a need to understand situational factors to align with and prioritize one or the other approach, rather than presume that an adaptive approach is always appropriate.

Next, we summarize this shared narrative and identify assumptions contrasted in the reviewed literature, shown in Figure 2.

Conventional approach. A conventional approach stems from Newtonian science grounded in reductionism and a logic of prediction and control for problem-solving through a "neat,

sequential and contained solution" (Putnam-Walkerly, 2017: 9). Goals—assumed to be clear, stable, and agreed-on—are to be achieved in a stable context defined as if reality is divisible, separable, simple, and finite (Birney, 2021). Incremental change focuses on discrete policies and programs to efficiently and effectively accomplish the goals. Each policy or program exists primarily to achieve one goal; the longevity and funding of the intervention exist until goal accomplishment (Renger et al., 2020). As cited in sustainability literature about the agricultural sector, "a largely quantitative, linear approach and a predominant interest in mainstreaming technological solutions" has led to a "rather narrow focus on incremental actions and adjustment approaches" (Silici et al., 2022: 190).

Traditional evaluation requires looking at anticipated outcomes within predetermined timeframes. Designs and methodologies orient around a linear depiction of causality (cause-effect logic models) in which interventions cause changes assessed in the evaluation (Hargreaves, 2018). Evaluating interventions requires and reinforces clear foci and standardization of the intervention and the evaluation design, methodology, and reporting. Evaluation underemphasizes outcomes not predicted by the model and instead focuses on validating the preconceived model of causality (Morell, 2018). Useful evaluations inform which interventions to replicate, scale up, or transfer to new contexts to accomplish specific goals or outcomes.

Systemic approach. A systemic approach reduces the possibility of neatly bounding and empirically explaining a problem and instead centers uncertainty and interconnectedness. In sustainability, this involves a shift from "techno-fixes" to "complex multi-dimensional issues" shaped by the social and political behaviors of human actors within institutional dynamics (Abson et al., 2017: 31), and from "static to dynamic thinking" and recognizing that systems are "greater than the sum of parts" (Willis et al., 2014: 2). The "classical paradigm of science and engineering" is rejected for social-ecological problems and systems. Goals cannot be clearly defined and agreed-on in advance. Without a neatly bounded problem and clear goals, foci shift to influencing "progress in changing the drivers of systems, the behaviours of system actors, and the overall behaviour of the system" (Cabaj, 2018: 10) within complex situations and systems rather than on completing a bounded change process. To do this, the literature emphasized identifying leverage points or areas to intervene while recognizing that interventions "work in highly context-specific ways" (Hendy and Barlow, 2012: 1). For Nakaima et al. (2013: 5), the question is not simply: "Does intervention 'A' work? Instead, the question is: How best does the 'ecology of services' offered by different organizations work to make a difference on unmet needs and quality of services?"

This requires ongoing, collaborative, engaged research and evaluation processes. When dealing with complex issues, change can be unpredictable; evaluation needs to be flexible and adapt constantly (Eoyang and Oakden, 2016). Rather than situating researchers and evaluators as external experts with answers, this requires the collaborative engagement of practitioners, civil society, and communities. In sustainability, Abson et al. (2017: 36) state that people "who function in the midst of sustainability challenges are often more aware and familiar with the level of complexity than academic scholars and hence can play a vital role in problem identification and solution development." Because interventions and contexts interact unpredictably, intervention and evaluation designs must be flexible (English et al., 2020: 13), eliminating what Eoyang and Oakden (2016) call a "false distinction" between evaluation and learning. Evaluation cannot happen post hoc; it needs to run parallel to intervention and serve the primary purpose of informing action as it unfolds (Atkinson et al., 2021). Post hoc evaluation of

merit and value is superseded by learning to guide a continuous adaptation process (Eoyang and Oakden, 2016). Merging the evaluation and learning processes increases the system's adaptation capacity (Nieminen and Hyytinen, 2015).

Causality is more tangled, and attributing impact to an intervention is difficult (Junge et al., 2020; Picciotto, 2020); a more nuanced view of causality can be cyclical, recursive, multidirectional, and unpredictable. Unintended consequences are not viewed as secondary in importance but as outcomes that deepen understandings around the complexity of causes and effects in dynamic, real-world settings (Hargreaves, 2018). Evaluations inform processes of interrogation and (re)adjustment. Monitoring, timely feedback, and adaptive management gain greater importance and become intrinsic components of the implementation plan. New interventions are treated as hypotheses to be piloted and tested. Evidence is conditional, bound by time and context. Good-enough evidence informs learning and action, with a look to the future. "Foresight" becomes a more systematic and participatory way of analyzing multiple possible futures and building a medium- to long-term vision for the future (Nieminen and Hyytinen, 2015).

Ten shared leverage areas with least attention to power and resources

A third set of results addresses 10 cross-cutting leverage areas to influence change and transformation. Some references used the term leverage point, while others identified lenses on change, drivers or factors to affect change, levels of intervention, and principles. We use leverage area to refer to the potential foci for change. In each field, there were popular frameworks and guides, but no references drew on interdisciplinary, cross-field scholarship. We coded leverage areas identified within references, clustered and synthesized these leverage areas across fields, and ordered each by how often it was discussed in the literature. We list those coded most often, in descending order, and describe each in Table 4.

These 10 shared leverage areas include the following: (1) responsiveness to context; (2) connections, relationships, networks, and alignment; (3) peoples' leadership and roles; (4) mental models, narratives, and memes; (5) shared vision; (6) practices and behaviors; (7) innovations and prototypes; (8) policies, laws, and incentives; (9) power; and (10) resources. Together, these leverage areas hold promise for developing and testing theories of where and how to influence change. Despite the central focus on systems change and transformation in the reviewed literature, few references provided detailed theories, frameworks, or models of change and transformation (for an exception, see Waddell et al., 2015). Within evaluation, the focus is predominantly placed on how evaluation can support systems change and transformation processes rather than on these processes themselves, except for efforts to develop theories of transformation (Patton, 2020). To us, this suggests an opportunity to develop theoretically grounded and empirically verifiable leverage areas as the science and practice of systems change and transformation evolves.

With regards to the most and least discussed leverage areas, it is surprising to see such little attention on power and resources. If systems change and transformation require negotiation (i.e. cannot be realized by any single actor) (Andrachuk and Armitage, 2015; Lee and Waddock, 2021; Van Bruggen et al., 2019), then power must be addressed. The creation of fundamentally new structures, through which new systems emerge, implies power changes that upend the current system(s). Many of the leverage areas, such as people's roles, practices, behaviors, and innovations, undoubtedly require shifts in resources, including funding, time, and

Table 4. Ten leverage areas for systems change and transformation.

Area	Description
Responsiveness to context	• Examine the surrounding context and user needs to identify ways to modify initiatives • Consider path dependency, attractors, and thresholds
Connections, relationships, networks, and alignment	• Strengthen connections and coherence between actors across sectors, roles, hierarchical levels, and between multiple initiatives • Cultivate social networks and social capital to align vision, exchange information and learning, and coordinate actions • Connect factors and explore feedback loops across time and scale
People: leadership and roles	• Consider types of leadership, coordination, and the roles of people in a change effort
Mental models, narratives, and memes	• Use language to bring multiple viewpoints together into a shared narrative • Shift how people understand what's happening or what's possible • Shift values, beliefs, and patterns of social and individual behavior
Shared vision	• Develop a shared vision or purpose to offer direction across barriers to work together over the long term, and in the face of uncertainty and diverse perspectives
Practices and behaviors	• Use guidelines, training, incentives, and other ways to influence individual and group behavior • Consider the degree of specification versus openness or flexibility
Innovations and prototypes	• Try out new things as prototypes and then encourage the diffusion of knowledge or use for those that work • Invest in research and development, especially in technology
Policies, laws, and incentives	• Change policies, laws, rules, and official practices to influence political and legal institutions and official practices by
Power	• Consider the type and degree of potential influence or agency that different agents have and the redistribution of this influence
Resources	• Reallocate funding, personnel, project management, time, investment, and other material conditions, including knowledge

knowledge, among others. Yet, resources were rarely discussed in the reviewed literature as a way to influence change. Connecting to the first finding, the lack of normative debate and presumption of universality may contribute to the minimal exploration and emphasis on shifts in power and resources as strategies to influence change and transformation.

Expanded role for MEL in change process poses challenges and opportunities

Results support an expanded role for MEL embedded throughout a systems change or transformation process. Discussions of MEL related to five phases or activities, captured in Table 5: critically diagnose issues and systems, iteratively design and implement interventions, measure and monitor, assess value and impact, and learn and adapt. We identified these phases during our analysis as references discussed MEL in relation to each of these, but did not present them in an organized manner. For each phase, we identify both challenges and opportunities. Some authors contend that learning must replace evaluation or that evaluation must be reimagined. We believe that is it important to ground these claims in a more thorough examination of the considerations at each phase. Note that results for this research question draw primarily from evaluation and philanthropy, as health, sustainability, and organizational change minimally address implications for MEL.

Table 5. Expanded role of evaluation in five phases of a change process.

Phase	Challenges	Opportunities
Critically diagnose issues and systems	• Nature of wicked problems—complex, uncertain, ongoing change • Natural, human, social, and financial interconnections • Assumptions about change based on colonizing, capitalist paradigms	• Shifts from problem-solving to social navigation and adaptive management • Systems modeling and mapping to analyze wider dynamics and processes • Early and ongoing consideration of what the 'problem' and 'success' mean
Iteratively design and implement	• Limits to prescribing then controlling change • Insufficiency of discrete projects, programs, and policies or incremental change	• Develop a research-based model or theory of change • Blend "what's working" and research from elsewhere with local, context/system-specific knowledge • Shift focus to leverage points or areas of influence • Capture spontaneously emerging activities and their outcomes
Measure and monitor	• Risks of measurement and limits to validity of instrumentation • Limitations of any one or several indicators to capture complexity • Lack of clarity and agreement on what to measure or monitor • New ideas for what to measure and monitor continuously emerge	• Focus on process and "the journey" • Iterate between model building and data collection • Track across levels, time, and perspectives, types of leverage, unintended or unexpected changes, and barriers or constraints to change • Build consistency in what and how data is tracked; data infrastructure • Bundle discrete indicators and criteria; present and interpret in context • Consider what to measure and how to use participatory and transparent approaches • Anticipate consequences of measures
Assess impact and value	• Design challenges due to lack of stability; attribution less relevant and possible given nested and interconnectedness; context specificity; small or no effect size at population and large scales • Plurality of values and perspectives and types of evidence pose challenges for composite scores	• Lots of new methods, such as process tracing, contribution analysis, outcome harvesting, social network analysis, and comparative case analysis • Need frameworks and processes (e.g. rubrics) that bring different types of evidence or data together with different values or perspectives
Learn and adapt	• Lack of culture and infrastructure to support ongoing evaluation and evidence use amid uncertainty and complexity • Pressure for results-based accountability and management poses risks for failure and dis-incentivizes experimentation	• Build evaluation culture with shared norms and a common language • Provide ongoing feedback and facilitate learning, formally and informally • Foster immediate experimentation and adaptation and broader learning and dissemination across contexts

Critically diagnose issues and systems. The nature of wicked problems was cited as a primary challenge for MEL. Wicked problems are ridden with "interdependent factors," "unpredictable dynamics," and "no feasible solutions" (Uitto et al., 2019: 127). Challenges are amplified by the recognition of natural, human, social, and financial capital interconnections (Uitto et al., 2019); the local to the global scale of issues (Ofir and Rugg, 2021); the need to consider ways to move global trajectories away from trends (Olsson et al., 2017); and reckoning with underlying assumptions about problems, change, and success embedded in colonizing, capitalist paradigms (Australian Centre for Social Innovation, 2019). A primary opportunity for MEL is to shift from a linear problem-solving orientation to continual social navigation and adaptive management (Ofir and Rugg, 2021). Systems mapping and modeling offer ways to analyze broader dynamics and processes (Scott and Pringle, 2018). Defining problems and success must be ongoing and continually revised as the situation and knowledge change over time.

Iteratively design and implement. Systems change and transformation pose challenges for MEL, in that they require iterative design and implementation. Change cannot be prescribed and accomplished (Junge et al., 2020). Discrete projects, programs, and policies may bring incremental change but are insufficient for transformation. Opportunities include "seeing evaluation as part of the change process (supporting or even amplifying it in intervention mode, and/or support learning and debate)" (Junge et al., 2020; Molas-Gallart et al., 2021). When evaluation shifts from testing an intervention model that addresses a pre-defined need or problem to embedded, ongoing MEL, actors can better "figure out what's really needed" (Walker, 2017). This may include developing a theory of change or transformation that brings in established and new research to inform local context or system-specific knowledge (Molas-Gallart et al., 2021). Development of a theory of change should include the "uncomfortable process" of reflecting on assumptions and distinguishing between a theory of change and a theory of inequities (Solar and Frenz, 2017: 10). Interventions should work on multiple levels and address leverage points, especially deeper ones (Atkinson et al., 2021; Uitto et al., 2019). Systematic reflections should not narrowly focus on implementation and outputs but more broadly explore dynamics and processes to identify new opportunities and designs (Loorbach et al., 2017; Norman, 2021).

Measure and monitor. There are numerous challenges to measuring and monitoring and some opportunities. As circumstances shift, new ideas for what to measure and monitor continuously emerge, alongside a need to "make sense of their relevance" (Junge et al., 2020).

A primary challenge stems from measurement risks and limits to instrumentation validity, given the focus on multiple factors and actors. Measurement is inherently reductionist and time-bound; this poses limitations for any or several indicators to capture complexity, especially at global scales (Patton, 2019) and across multiple interdependent factors (Garnett, 2014). Limited theoretical and empirical research addresses how to change or transform systems, and little clarity or agreement exists among experts regarding *how* to conceptualize and operationalize what to measure or how to best develop measures. Resilience scholars, for example, disagree about the value of measuring individual elements of a socioecological system. While there is some shared interest in understanding the resilience of a system as a whole and resilience as a mind-set, established measures do not yet exist on either front (Andrachuk and Armitage, 2015).

Despite these challenges, there are opportunities for measurement and monitoring. With respect to orientation, several authors emphasize focusing on "the journey" (Willis et al., 2014: 12). Process considerations include iterating between model building and data collection, and ongoing consideration of tradeoffs between rigor and flexibility (Morell, 2018); collaborative and transparent processes to define what to measure and how (Burch et al., 2014); developing sound data tracking and management infrastructure to help ensure consistency (Perla et al., 2013); and how who gets monitored might themselves help identify and address inequities (Nakaima et al., 2013: 9). Discernment regarding what to measure or monitor should be strategic and reflexive (Rijswijk et al., 2015). Ideas about what to measure and monitor include:

- tracking across levels (e.g. individual, organizational, environmental), time (e.g. retrospective, real-time, future scenarios), and perspective (e.g. those who join earlier or later, roles) (Atkins et al., 2017; Norman, 2021)
- tracking types of leverage (e.g. degrees of convergence and divergence of mental models) (Evans et al., 2015)
- mechanisms that drive or explain variations in practice or behavior (Atkins et al., 2017)
- dynamics of change over time (Koleros et al., 2016)
- unintended or unexpected changes (Morell, 2018; Uitto et al., 2019)

Assess value and impact. A recurring challenge for evaluators/evaluation is the infeasibility of established designs. Design challenges stem from multiple actors and factors, making cause–effect relationships difficult to discern and attribution inappropriate (Molas-Gallart et al., 2021; Stachowiak et al., 2020). Nestedness, interconnections, and time horizons add to this difficulty (Burch et al., 2014). "Looking beyond individual investments in scope, scale, and time . . . is not what evaluations typically do, nor do we have the mandate or types of evaluative tools for this" (Uitto et al., 2019: 127), foreshadowing a rise in portfolio based evaluation (UNDP, 2022).

> Constant movement creates the specific challenge of not being able to look back over a period of stability in order to measure the change caused by an intervention. Therefore, emergence, coupled with the long-term nature of the change that is being sought and its scale is the key challenge for evaluators seeking to assess outcomes and impacts (Junge et al., 2020: 232)

Normative questions exist around defining impact, success, value, and progress in contexts of multiple values and perspectives. In sustainability, questions about appropriate time horizons (Norman, 2021) and "fundamental change as opposed to surface level change" (Narayanan and Adams, 2016: 348) are amplified by the need for longer timeframes to understand system dynamics (Loorbach et al., 2017). Investments tied to electoral cycles are a constraint pattern (Burch et al., 2014), and there is no consensus around what social innovation means, or how to distinguish social innovation success (Antadze and Westley, 2012). Generalizability limitations are another factor, a topic long debated within the evaluation field and renewed in relation to the local and unique characteristics of any complex adaptive system (Hendy and Barlow, 2012),. Large effect sizes are unlikely to be seen for population-level interventions (Jones et al., 2016), though effect sizes are less important than durability and persistence over time, in non-linear development processes (Silici et al., 2022). Despite these challenges,

"assessments of progress and recalibration are important" (Siegel et al., 2015: 18). Determinations of value must also mitigate against tendencies to "report on success stories" and risks of bias if "values do not recognize the importance of diversity and the inextricable interconnections between the well-being of humanity and nature" (Atkinson et al., 2021: 136).

Opportunities widely discussed center on new methodologies that address the contribution and identification of patterns with transferability, account for synergies and unintended consequences, and look at durability and persistence. Methods identified as promising in the reviewed literature include process tracing, modus operandi, ripple effect mapping, contribution analysis, qualitative comparison analysis, outcome harvesting, social network analysis, group concept mapping, causal loop diagramming, agent-based modeling, systems mapping, driver diagrams, comparative cases to examine how structures operate, and innovative economic methods (see Beer, 2017; Nieminen and Hyytinen, 2015). Another opportunity relates to frameworks and processes that bring different types of evidence, data, or metrics together with other values or perspectives to support sense-making when analysis units vary across epistemologies (Jones et al., 2016: 16). It is important to consider relationships when outcomes cut across sectors (Garnett, 2014; Silici et al., 2022) and to recognize how degrees of consensus and uncertainty shape evidence bases (Silici et al., 2022). Processes may include participatory approaches and mutual learning to overcome presumed impartiality. One might shift from "the number or quality of relationships in a systems leadership intervention" to "how leaders weave these together to create the relational cradle" (Atkinson et al., 2021: 134). Or from "what is valued as productive and efficient in agriculture" to inter- and intra-generational equity (including ecological sustainability) embedded into the assessment of outcomes (Nakaima et al., 2013: 9).

Learn and adapt. At organizational, sector, and cross-sector levels, culture and infrastructure supports are needed for ongoing evaluation and evidence use, amid uncertainty and complexity. Pressure for results-based accountability poses risks for failure and dis-incentivizes experimentation (Van Ongevalle et al., 2014). Even when evaluations of systems change and transformation do occur, it is not apparent how to synthesize evidence and transfer knowledge. This is, in part, due to emphasis on "match[ing] our methods to each situation at hand" (Rockefeller Philanthropy Advisors, 2019: 5) and the use of contextually rich, case-based designs (Chino, 2012; Hubeau et al., 2017). Reviewed literature underscores the importance of a repository of approaches and lessons (Perla et al., 2013) and fora to exchange practices and lessons learned (Nakaima et al., 2013).

Opportunities for the role of MEL in promoting learning and adaptation begin with building an evaluation culture with shared norms and a common language (Norman, 2021; Sturmberg et al., 2012). This involves valuing data and making "data collection, analysis, reporting, and learning . . . routine operations" and "not just the work of one person or unit" (Baptiste and Iese, 2019: 82). Others advise planning for learning linked with technical assistance (Scott and Pringle, 2018), fostering humility, and embracing discomfort with the unknown. A second opportunity relates to the provision of ongoing feedback that facilitates learning. Forms of learning include formal and informal (Scott and Pringle, 2018); single, double, and triple loop (Baptiste and Iese, 2019); quick and efficient feedback to inform adaptation (Atkins et al., 2017); and broader learning and dissemination across contexts. Ideas to build formal learning systems include coordinating data collection across sites or initiatives (Nakaima et al., 2013) and using distributed learning technologies, data management systems,

and evidence-based learning tools and techniques (Scott and Pringle, 2018). A third opportunity requires shifting our conception of evaluation quality from products (i.e. plans, reports, deliverables) to process. This, in turn, requires norms for the evaluation process that could include careful boundary setting with attention to context and interconnectivity (Naidoo, 2022), mapping throughout a change initiative to show system and navigation (Hussey et al., 2021), and enabling transformation through evaluation by, for example, opening space for evaluative conversations (Naidoo, 2022).

Bridging findings with ongoing and recent evaluation debates

The first finding highlights a striking absence of attention on and debates about normative issues. Within reviewed references, systems change and transformation were often positioned as widely shared ideals with universal appeal. A rationale, if any was given, was grounded in the urgency and complexity of large-scale problems and environmental sustainability as fundamental to the survival of humanity. While this finding captures a cross-field pattern, it misses ongoing discussion within the evaluation literature regarding whose values, perspectives, and worldviews underpin and should drive change and transformation efforts (Schwandt and Gates, 2021; Mertens, 2010). In addition, this contrasts with definitions found in evaluation literature, where systems change and transformation are not only described in structural or functional terms but are also framed as normative processes. Authors such as Patton, (2019, 2021), Preskill et al. (2014), and Brousselle and McDavid (2020) emphasize the importance of values, power, equity, and worldview in shaping the direction and meaning of transformation. Definitions foreground contested perspectives, emergent outcomes, and the evaluator's role in navigating complexity, challenging transformation as a universally desirable or inherently positive goal.

The second finding features a persistent and sometimes simplistic narrative of moving from conventional approaches to program and policy design, implementation, and evaluation to systems change and transformation as a systemic alternative. This narrative continues to pervade much of evaluation literature even going so far as a paradigm shift (Patton, 2023; Picciotto, 2020). However, the idea is not to expand and enhance evaluators' toolkits (Gates et al., 2021) while continuing to use and refine fundamental and established evaluation concepts and methods, such as the expansion of theory-based evaluation to guide and assess intervening within socioecological systems (Miyaguchi, 2022) and collaboratively generating and using learning questions to navigate uncertainty (Bertermann and Coffman, 2024). In addition, there are a plurality of ways of thinking and working informed by complexity and geared toward systems change and transformation rather than a single, agreed-on systemic approach (Gates et al., 2021; Barbrook-Johnson et al., 2021).

The third finding identifies 10 common leverage areas across the five included domains of evaluation, health, organizational change, sustainability, and philanthropy. There have been numerous organizational efforts to propose frameworks, such as the popular *Water of Systems Change* by Kania et al. (2018) and field-specific models, including five principles in large-scale change in public health (Best et al., 2012), three layers of processes of change in social-ecological systems (Barnes et al., 2017) and four levels of potential in changing systems (Birney, 2021). Building on these frameworks, we see an opportunity for interdisciplinary, integrative research on leverage areas for systems change and transformation. The 10 areas we list offer a starting point to be expanded by incorporating recent literature and a more robust

comparison of organizational and field-specific models. In addition, it is important to note that leverage areas is not the only way to frame and carry out efforts to change systems. For example, Lynn and Coffman (2024) contrast emergent and systems dynamic mental models in philanthropy and identify leverage areas only within the latter approach.

Finally, some of the challenges for evaluation identified in this review have been or are being addressed. MEL is increasingly expected to inform strategy, surface assumptions, track emergent patterns, and support adaptive decision-making. There have been developments around theories of change and transformation, with Rogers (2024) providing guidance on developing a theory of change for interventions into complex systems; Patton and Richardson (2024) showcasing an example theory of transformation for a global network shifting the food systems; and Koleros et al. (2016) continuing to develop an actor-based theory of change approach and building the community of practice around theories of change (Koleros et al., 2024). What were at first fringe methodologies now are prominently used, including evaluative rubrics, outcome harvesting, and contribution analysis, among others. In sync with the increased use of participatory approaches, there are efforts to pluralize notions of evidence quality, such as inclusive rigor, within systems change efforts (Apgar et al., 2024).

These connections show some ways review findings pose relevance, while leaving a fuller examination of findings in light of recent scholarship open for reader consideration.

Implications and future directions by role

Review results provide support for shifts underway within the evaluation field among funders working collaboratively, evaluators expanding their skillsets, and intermediary agencies and networks facilitating transdisciplinary exchanges.

Funders and commissioners

This review underscores a need to rethink power relations to shape new roles for funders and commissioners. As, for instance, Picciotto (2020: 73) makes the case that "break[ing] free from the tyranny of market forces . . . will require reformed knowledge regimes, new evaluation governance systems, and diversification of funding sources. In philanthropy, private investment, and international aid and development, notable efforts to shift the role of funders have steadily grown since the time of this review. These include trust-based philanthropy (Trust-based Philanthropy Project, 2021), Rockefeller Philanthropy Advisors' Shifting Systems Initiative, and systemic investing (Daggers et al., 2023; Hofstetter, 2025), and the Equitable Evaluation Initiative (Dean-Coffey and Coné, 2023). Shared emphases of these efforts include the importance of practical, co-designed evaluation terms of reference and scopes of work that address experimentation, learning, and adaptation. These contractual documents (and the mental models they capture and call for) should reflect a shift beyond narrowly conceived grants, projects, goals, and indicators to allow for iterative, collaborative change processes. Work on systemic investing aims to shift the deployment of capital from independent, targeted efforts assessed for financial returns to collaborative processes among multiple types of funders, organizational partners, and communities to shift system conditions and outcomes (Daggers et al., 2023; Hofstetter, 2025).

Evaluators and evaluation educators

Another implication of this review is the necessity to shift from individual evaluators to multidisciplinary teams, posing implications for evaluator competencies and evaluation education. The breadth and variety of theoretical and methodological knowledge and skills needed for an adaptive approach to MEL suggest the necessity of working on multidisciplinary teams where general knowledge and specialization consistently converge (Ofir and Rugg, 2021; Picciotto, 2020; Uitto et al., 2019). Mix "traditional evaluative skills with innovative ones," which "few individuals or even organizations have" (Uitto et al., 2019: 127). Since this review, a wealth of guidelines for systems change and evaluation have been published outside academic scholarship. Similarly, there are well-established courses, workshops, and trainings in the systems and complexity sciences, such as those offered by the International Society for the Systems Sciences (https://www.isss.org/), the Open University's Systems Thinking in Practice program (https://www.open.ac.uk/courses/choose/systemsthinking), and the Santa Fe Institute (santafe.edu). Still, we see an opportunity to incorporate systems thinking, methods, and practices into higher education (Van den Berg et al., 2019), especially in evaluation education programs.

Transdisciplinary research communities

The review highlights the importance of intermediary agencies and networks facilitating transdisciplinary exchanges. Fortunately, there are numerous examples of prominent efforts which bring together funders across sectors, practitioners, and evaluators. These include:

- United Nations Development Programme (UNDP)'s Strategic Innovation Unit and the Unstuck initiative bring together governments, communities, and partners to explore how portfolio and systems approaches can drive transformative change in public systems.
- Co-Impact, a global philanthropic collaborative, explicitly supports systems change by funding and co-creating large-scale, locally grounded strategies, often integrating evaluation and learning components with government partners.
- The Causal Pathways Initiative helps funders and evaluators explore alternative approaches to causality that move beyond experimental designs, aiming to better capture complex and systemic effects of social interventions.
- The Inclusive Rigor Co-lab, hosted by the Institute of Development Studies (IDS), fosters collective inquiry into how peace-building and evaluation can be shaped by inclusive, power-aware, and context-sensitive methodologies.

In addition to mutual learning and networking, these initiatives have led the development of case examples to illustrate and diffuse innovative approaches. In tandem, we see a need for comparative case analyses and the use of shared theoretical frameworks. As mentioned, there are a plethora of systems change models and guidance, pointing to an opportunity for a crosswalk that compares and contrasts to build shared language and learning. In particular, drawing on this review, we see the potential for extending the 10 cross-leverage areas to define each further, theorize and test the relationships, and delineate strategies and interventions.

Conclusion

As global systems face mounting pressures—from climate instability to deepening social inequities—the capacity to evaluate change systemically and critically is both timely and necessary. This conclusion synthesizes cross-cutting insights from the integrative review and identifies areas for continued inquiry to advance evaluation in support of systems change and transformation. The findings of this review—drawn from over a decade of academic and gray literature across evaluation, health, organizational change, sustainability, and philanthropy—point to persistent gaps and emerging opportunities that require concerted attention across fields and actors. This review further highlights a need to critically engage with the normative dimensions of systems change and transformation, confront power and resources as leverage areas, and reimagine the evaluative practices needed to navigate transformation. Review results provide support for shifts underway within the evaluation field among funders rethinking their roles to work collaboratively, evaluators expanding their skillsets and joining multidisciplinary teams, and intermediary agencies facilitating transdisciplinary exchanges to share learning and coordinate action within an evolving field and practice.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD


Emily F. Gates https://orcid.org/0000-0003-2352-5064

References

Abercrombie R, Harries E and Wharton R (2015) *Systems Change: A Guide to What It Is and How to Do It*. London: New Philanthropy Capital (NPC).

Abson DJ, Fischer J, Leventon J, et al. (2017) Leverage points for sustainability transformation. *Ambio* 46(1): 30–9.

Andrachuk M and Armitage D (2015) Understanding social-ecological change and transformation through community perceptions of system identity. *Ecology and Society* 20(4): 26.

Antadze N and Westley FR (2012) Impact metrics for social innovation: barriers or bridges to radical change? *Journal of Social Entrepreneurship* 3(2): 133–50.

Apgar M, Aston T, Snijder M, et al. (2024) Raising the bar: Improving how to assess evidence quality in evaluating systems-change efforts. *The Foundation Review* 16(2): 7–21.

Atkins D, Kilbourne AM and Shulkin D (2017) Moving from discovery to system-wide change: the role of research in a learning health care system. *Annual Review of Public Health* 38(1): 467–87.

Atkinson J, Lasbennes F and Nabarro D (2021) Reflecting on our times: valuing transformative leadership in real-world "living systems." *American Journal of Evaluation* 42(1): 130–8.

Gates EF (2016) Making sense of the emerging conversation in evaluation about systems thinking and complexity science. *Evaluation and Program Planning* 59: 62–73.

Gates EF (2018) Towards valuing with critical systems heuristics. *American Journal of Evaluation* 39(2): 201–220.

Gates EF,Walton M, Vidueira P, et al. (2021) Introducing systems- and complexity-informed evaluation. *New Directions for Evaluation* 2021: 13–25.

Baptiste L and Iese V (2019) The transformative agenda for evaluation in small island developing states: The Caribbean and the Pacific. In: Van den Berg R, Magro C and Mulder S (eds) *Evaluation for Transformational Change: Opportunities and Challenges for the Sustainable Development Goals*. Exeter: International Development Evaluation Association (IDEAS), 34–48.

Barbrook-Johnson P, Castellani B, Hills D, et al. (2021) Policy evaluation for a complex world: Practical methods and reflections from the UK centre for the evaluation of complexity across the Nexus. *Evaluation* 27(1): 4–17.

Barnes M, Bodin Ö, Guerrero AM, et al. (2017) The social structural foundations of adaptation and transformation in social–ecological systems. *Ecology and Society* 22(4): 16.

Beer T (2017) *Systems Change Evaluation Forum: Executive Summary Report. Center for Evaluation Innovation*. Available at: https://www.evaluationinnovation.org/wp-content/uploads/2017/05/final-systems-evaluation-forum.pdf (accessed 23 June 2025).

Bertermann K and Coffman J (2024) Embedding learning in systems change: A learning framework for testing uncertainties. *Center for Evaluation Innovation*. Available at: https://evaluationinnovation.org/wp-content/uploads/2024/12/v2-Client_CEI_Embedded-Learning-Brief.pdf (accessed 23 June 2025).

Best A, Greenhalgh T, Lewis S, et al. (2012) Large-system transformation in health care: A realist review. *The Milbank Quarterly* 90(3): 421–56.

Birney A (2021) How do we know where there is potential to intervene and leverage impact in a changing system? The practitioner's perspective. *Sustainability Science* 16(3): 749–65.

Braithwaite J, Churruca K, Long JC, et al. (2018) When complexity science meets implementation science: A theoretical and empirical analysis of systems change. *BMC Medicine* 16(1): 63.

Brousselle A and McDavid J (2020) Evaluators in the anthropocene. *Evaluation* 26(2): 190–204.

Burch S, Shaw A, Dale A, et al. (2014) Triggering transformative change: A development path approach to climate change response in communities. *Climate Policy* 14(4): 467–87.

Cabaj M (2018) *Evaluating Systems Change Results: An Inquiry Framework*. Available at: https://www.tamarackcommunity.ca/hubfs/Resources/Publications/Paper%20Evaluating%20Systems%20Change%20Results%20Mark%20Cabaj.pdf (accessed 23 June 2025).

Chino M (2012) Tribal capacity building as a complex adaptive system: New insights, new lessons learned. *International Public Health Journal* 4(3): 235–42.

Schwandt TA and Gates EF (2021) *Evaluating and Valuing in Social Research*. Guilford Press.

Cristancho S, Field E and Lingard L (2018) What is the state of complexity science in medical education research? *Medical Education* 53(1): 95–104.

Daggers J, Hannant A and Jay J (2023) Systemic investing for social change. *Stanford Social Innovation Review*. Available at: https://ssir.org/articles/entry/systemic_investing_for_social_change (accessed 23 June 2025).

Dean-Coffey J and Coné M (2023) *The Equitable Evaluation Framework: May 2023 Expansion*. Available at: https://www.equitableeval.org/framework (accessed 23 June 2025).

English M, Nzinga J, Irimu G, et al. (2020) Programme theory and linked intervention strategy for large—scale change to improve hospital care in a low and middle—income country—a study pre-protocol. *Wellcome Open Research* 5: 265.

Eoyang G and Oakden J (2016) Adaptive evaluation. *Emergence: Complexity and Organization* 18(3/4): 1–14.

Evans JM, Baker GR, Berta W, et al. (2015) Culture and cognition in health systems change. *Journal of Health Organization and Management* 29(7): 874–92.

Flyvbjerg B (2001) *Making Social Science Matter: Why Social Inquiry Fails and How It Can Succeed Again* (trans. S. Sampson). Cambridge: Cambridge University Press.

Garnett T (2014) Three perspectives on sustainable food security: Efficiency, demand restraint, food system transformation. What role for life cycle assessment? *Journal of Cleaner Production* 73: 10–8.

Grant MJ and Booth A (2009) A typology of reviews: An analysis of 14 review types and associated methodologies. *Health Information & Libraries Journal* 26(2): 91–108.

Hargreaves M (2018) *Leveraging Systemic Change: Evaluating What Works*. Chicago, IL: NORC at the University of Chicago.

Hendy J and Barlow J (2012) The role of the organizational champion in achieving health system change. *Social Science and Medicine* 74(3): 348–55.

Hofstetter D (2025) *The Limitations of Investing in Universal Solutions. Stanford Social Innovation Review*. Available at: https://ssir.org/articles/entry/systemic-investing-context-specific (accessed 23 June 2025).

Hubeau M, Marchand F, Coteur I, et al. (2017) A new agri-food systems sustainability approach to identify shared transformation pathways towards sustainability. *Ecological Economics* 131: 52–63.

Hussey AJ, Sibbald SL, Ferrone M, et al. (2021) Confronting complexity and supporting transformation through health systems mapping: a case study. *BMC Health Services Research* 21(1): 1146.

Jackson MC (2019) *Critical Systems Thinking and the Management of Complexity: Responsible Leadership for a Complex World*. Hoboken, NJ: Wiley Blackwell.

Jones L, Fraser A and Stewart E (2019) Exploring the neglected and hidden dimensions of large-scale healthcare change. *Sociology of Health and Illness* 41(7): 1221–35.

Jones M, Verity F, Warin M, et al. (2016) OPALesence: Epistemological pluralism in the evaluation of a systems-wide childhood obesity prevention program. *Evaluation* 22(1): 29–48.

Junge K, Cullen J and Iacopini G (2020) Using contribution analysis to evaluate large-scale transformation change processes. *Evaluation* 26(2): 227–45.

Kania J, Kramer M and Senge P (2018) *The Water of Systems Change*. Boston, MA: FSG. Available at: https://www.fsg.org/publications/water_of_systems_change (accessed 23 June 2025).

Koleros A, Adrien M-H and Tyrrell T (eds) (2024) *Theories of Change in Reality: Strengths, Limitations and Future Directions*. New York: Routledge.

Koleros A, Jupp D, Kirwan S, et al. (2016) Methodological considerations in evaluating long-term systems change: A case study from Eastern Nepal. *American Journal of Evaluation* 37(3): 364–80.

Lee JY and Waddock S (2021) How transformation catalysts take catalytic action. *Sustainability* 13(17): 9813.

Loorbach D, Frantzeskaki N and Avelino F (2017) Sustainability transitions research: Transforming science and practice for societal change. *Annual Review of Environment and Resources* 42: 599–626.

Lynn J and Coffman J (2024) Passing in the dark: Making visible philanthropy's hidden and conflicting mental models for systems change. *The Foundation Review* 16(1): 20–35.

Mertens DM (2010) Social transformation and evaluation. *Evaluation Journal of Australasia* 10(2): 3–10.

Miyaguchi T (2022) Importance and utilization of theory-based evaluations in the context of sustainable development and social-ecological systems. In: Uitto JI and Batra G (eds) *Transformational Change for People and the Planet: Evaluating Environment and Development*. Cham: Springer, 223–35. Available at: https://doi.org/10.1007/978-3-030-78853-7_15 (accessed 16 June 2025).

Molas-Gallart J, Boni A, Giachi S, et al. (2021) A formative approach to the evaluation of transformative innovation policies. *Research Evaluation* 30(4): rvab016.

Morell JA (2018) Systematic iteration between model and methodology: A proposed approach to evaluating unintended consequences. *Evaluation and Program Planning* 68: 243–52.

Naidoo IA (2022) Evaluation for transformational change: Learning from practice. In: Uitto JI and Batra G (eds) *Transformational Change for People and the Planet: Evaluating Environment and Development* (Sustainable Development Goals Series). Cham: Springer, 17–26. Available at: https://doi.org/10.1007/978-3-030-78853-7_2 (accessed 4 August 2022).

Nakaima A, Sridharan S and Gardner B (2013) Towards a performance measurement system for health equity in a local health integration network. *Evaluation and Program Planning* 36(1): 204–12.

Narayanan V and Adams CA (2016) Transformative change towards sustainability: The interaction between organisational discourses and organisational practices. *Accounting and Business Research* 47(3): 344–68.

Nieminen M and Hyytinen K (2015) Future-oriented impact assessment: Supporting strategic decision-making in complex socio-technical environments. *Evaluation* 21(4): 448–61.

Norman CD (2021) Supporting systems transformation through design-driven evaluation. *New Directions for Evaluation* 2021(170): 149–58.

Nyström ME, Höög E, Garvare R, et al. (2013) Change and learning strategies in large scale change programs: Describing the variation of strategies used in a health promotion program. *Journal of Organizational Change Management* 26(6): 1020–44.

Ofir Z and Rugg D (2021) American Journal of Evaluation section on international developments in evaluation: Transforming evaluation for times of global transformation. *American Journal of Evaluation* 42(1): 47–52.

Olsson A, Thunborg C, Björkman A, et al. (2020) A scoping review of complexity science in nursing. *Journal of Advanced Nursing* 76(8): 1961–76.

Olsson P, Moore M-L, Westley FR, et al. (2017) The concept of the Anthropocene as a game-changer: a new context for social innovation and transformations to sustainability. *Ecology and Society* 22(2): 31.

Palmatier RW, Houston MB and Hulland J (2017) Review articles: purpose, process, and structure. *Journal of the Academy of Marketing Science* 46(1): 15.

Palomo I, Locatelli B, Otero I, et al. (2021) Assessing nature-based solutions for transformative change. *One Earth* 4(5): 730–41.

Patton M and Richardson R (2024) A philanthropic theory of systems transformation for advancing equity in the polycrisis. *The Foundation Review* 16(2): 55–71.

Patton MQ (2019) Transformation to global sustainability: Implications for evaluation and evaluators. *New Directions for Evaluation* 2019(162): 103–17.

Patton MQ (2020) *Blue Marble Evaluation: Premises and Principles*. New York: The Guilford Press.

Patton MQ (2021) Evaluation criteria for evaluating transformation: Implications for the coronavirus pandemic and the global climate emergency. *American Journal of Evaluation* 42(1): 53–89.

Patton MQ (2023) Evaluating transformation means transforming evaluation. In: Edmondson B (ed) *Sustainability Transformations, Social Transitions and Environmental Accountabilities*. Cham: Springer International Publishing, 15–38. Available at: https://doi.org/10.1007/978-3-031-18268-6_2 (accessed 12 June 2025).

Perla RJ, Bradbury E and Gunther-Murphy C (2013) Large-scale improvement initiatives in healthcare: A scan of the literature. *Journal for Healthcare Quality* 35(1): 30–40.

Picciotto R (2020) Towards a complexity framework for transformative evaluation. *Journal of Multidisciplinary Evaluation* 16(35): 54–76.

Preskill H, Parkhurst M and Juster J (2014) *Guide to Evaluating Collective Impact*. Cambridge, MA: FSG. Available at: https://www.fsg.org/resource/guide-evaluating-collective-impact/ (accessed 23 June 2025).

Putnam-Walkerly K (2017) *The Role of Philanthropy in Systems Change: Findings and Lessons from a Field Scan of Systems Change and Policy Advocacy in the Philanthropic Field*. Cleveland, OH: Putnam Consulting Group.

Renger R, Foltysova J, Renger J, et al. (2020) Comparing and contrasting a program versus system approach to evaluation: The example of a cardiac care system. *Canadian Journal of Program Evaluation* 35(2): 210–27.

Rijswijk K, Bewsell D, Small B, et al. (2015) Reflexive monitoring in New Zealand: Evaluation lessons in supporting transformative change. *Evaluation Journal of Australasia* 15(4): 38–43.

Rockefeller Philanthropy Advisors (2019) *Assessing Systems Change: A Funders' Workshop Report*. Workshop report, October. New York: Rockefeller Philanthropy Advisors.

Rogers P (2024) Developing, representing, and using theories of change for interventions in complex systems. In: Koleros A, Adrien M-H and Tyrrell T (eds) *Theories of Change in Reality: Strengths, Limitations and Future Directions*. New York: Routledge, 94–102.

Rusoja E, Haynie D, Sievers J, et al. (2018) Thinking about complexity in health: A systematic review of the key systems thinking and complexity ideas in health. *Journal of Evaluation in Clinical Practice* 24(3): 600–6.

Scally CP, Lo L, Pettit KLS, et al. (2020) *Driving Systems Change Forward: Leveraging Multisite, Cross-sector Initiatives to Change Systems, Advance Racial Equity, and Shift Power*. Washington, DC: Federal Reserve Bank of San Francisco and Urban Institute.

Scott KA and Pringle J (2018) The power of the frame: Systems transformation framework for health care leaders. *Nursing Administration Quarterly* 42(1): 4–14.

Siegel B, Winey D and Kornetsky A (2015) *Pathways to System Change: The Design of Multisite, Cross-sector Initiatives*. Working paper. San Francisco, CA: Federal Reserve Bank of San Francisco and Mt. Auburn Associates, Inc.

Silici L, Knox J, Rowe A, et al. (2022) Evaluating transformational adaptation in smallholder farming: Insights from an evidence review. In: Uitto JI and Batra G (eds) *Transformational Change for People and the Planet*. Sustainable Development Goals Series. Cham: Springer, 187–202. Available at: https://link.springer.com/10.1007/978-3-030-78853-7_13 (accessed 18 January 2025).

Solar O and Frenz P (2017) Lessons from Chile's use of system-level theory of change to implement a policy redesign process to address health inequities. *New Directions for Evaluation* 2017(154): 101–13.

Stachowiak S, Lynn J and Akey T (2020) Finding the impact: Methods for assessing the contribution of collective impact to systems and population change in a multi-site study. *New Directions for Evaluation* 2020(165): 29–44.

Sturmberg JP, Halloran DMO and Martin CM (2012) Understanding health system reform—A complex adaptive systems perspective. *Journal of Evaluation in Clinical Practice* 18(1): 202–8.

The Australian Centre for Social Innovation (2019) *Philanthropy, Systems and Change*. Adelaide, SA, Australia: The Australian Centre for Social Innovation.

Trust-based Philanthropy Project (2021) *Using Trust-based Values to Guide Your Philanthropy's Culture, Structures, Leadership, and Practice*. Available at: https://static1.squarespace.com/static/607452f8ad01dc4dd54fc41f/t/61606874440b79448fb082c3/1633708148997/TBp+in+4D_Oct2021.pdf (accessed 23 June 2025).

Uitto JI, Puri J, Williams A, et al. (2019) Evaluating transformational change — Lessons from international environmental funds. In: Van den Berg R, Magro C and Mulder S (eds) *Evaluation for Transformational Change: Opportunities and Challenges for the Sustainable Development Goals*. Exeter: International Development Evaluation Association (IDEAS), 105–30.

Ulrich W (1983) *Critical Heuristics of Social Planning: A New Approach to Practical Philosophy*. Geneva: Swiss National Science Foundation.

United Nations Development Programme (UNDP) (2022) *System Change: A Guidebook for Adopting Portfolio Approaches*. Available at: https://unstuck.systems/system-approaches-toolkit/ (accessed 23 June 2025).

Van Bruggen A, Nikolic I and Kwakkel J (2019) Modeling with stakeholders for transformative change. *Sustainability* 11(3): 825.

Van den Berg R, Magro C and Mulder S (2019) *Evaluation for Transformational Change: Opportunities and Challenges for the Sustainable Development Goals*. Exeter: International Development Evaluation Association (IDEAS).

Van Ongevalle J, Huyse H and Van Petegem P (2014) Dealing with complexity through actor-focused planning, monitoring and evaluation (PME). *Evaluation* 20(4): 447–66.

Waddell S, Waddock S, Cornell S, et al. (2015) Large system change: An emerging field of transformation and transitions. *Journal of Corporate Citizenship* 58: 5–30.

Waddock S and Waddell S (2021) Transformation catalysts: Weaving transformational change for a flourishing world for all. *CADMUS* 4(4): 18.

Walker JC (2017) Solving the world's biggest problems: Better philanthropy through systems change. *Stanford Social Innovation Review*. Available at: https://ssir.org/articles/entry/better_philanthropy_through_systems_change (accessed 23 June 2025).

Westley F, Olsson P, Folke C, et al. (2011) Tipping toward sustainability: Emerging pathways of transformation. *AMBIO* 40(7): 762–80.

Whittemore R and Knafl K (2005) The integrative review: Updated methodology. *Journal of Advanced Nursing* 52(5): 546–53.

Willis CD, Best A, Riley B, et al. (2014) Systems thinking for transformational change in health. *Evidence and Policy: A Journal of Research, Debate and Practice* 10(1): 113–26.

Emily F. Gates is an Associate Professor of Evaluation at Boston College. Her research examines systems approaches to evaluation, including the role of values and equity. She holds a PhD from the University of Illinois at Urbana-Champaign.

Pablo Vidueira is an Assistant Professor of Evaluation and Food Systems at the Universidad Pontificia Comillas and Director of Evaluation at the Global Alliance for the Future of Food. He is an Agricultural and Biosystems Engineer, MSc. in Rural Development and Sustainable Management, and PhD. in Program Evaluation.

Maya Komakhidze is a PhD student in Human Development, Learning, and Teaching at Harvard University. She received her master's degree from Boston College in Applied Statistics and Psychometrics and previously worked as an evaluator and applied researcher in the fields of education and social development.

Charlotte Aldrich is the Assistant Research Director for TIMSS Mathematics and Special Projects at the TIMSS & PIRLS International Study Center at Boston College. Her research evaluates international education systems. She holds a PhD in Measurement, Evaluation, Statistics, and Assessment from Boston College.

Clara Shim is a Research Analyst at American Student Assistance. Her research evaluates educational policies and practices in Korea. She holds a PhD in Measurement, Evaluation, Statistics, and Assessment from Boston College.