

In the Making

Generative Evaluation: Learning With and For Living Systems

Malika Virah-Sawmy^a, Pierre Golbach^b, Christina Tewes-Gradl^c

Endeva GmbH, Berlin

^a*malikavs@gmail.com*

^b*p.golbach@endeva.org*

^c*c.gradl@endeva.org*

Abstract

Evaluation has so much potential for learning, co-creation, and collaboration around more inclusive and regenerative programs and, ultimately, systems. Yet, currently, evaluations are often experienced as controlling and frustrating add-ons. Inspired by developmental evaluation, we have experimented with generative methods to make evaluation a space for stakeholders to come together, re-connect, re-refresh the purpose of collaboration and re-view the design. We see that such a generative approach to evaluation can help to (1) express and co-construct systemic issues that often remain hidden and destructive otherwise, (2) create connection and energy for collaboration and (3) open a world of possibilities to learn from the past and adjust for the future. Organizations that use a more generative approach can hope to change the mindset within to more learning, collaboration, and courageous experimentation.

Keywords

evaluation, systems change, learning, generativity, emergence, transformation

Introduction

Evaluation matters. How success is defined and measured drives how interventions are designed and implemented. If evaluation strives for expertise, judgement, and measurement of the past, interventions will be designed for confirmation, certainty and measurability.

Can we evaluate instead in a more warm, organic, and wide way in order to enable more daring, complex, and exploratory interventions? These kinds of interventions are needed for the transformative changes we seek in many domains, from energy to food to health. This is where Generative Evaluation is coming from: the desire to learn collectively from complex interventions in order to chart our way forward, towards more regenerative and inclusive societies and systems. We understand Generative Evaluation as a set of practices that enables users to learn with and from the living system. It starts from the observation that there are tensions between any project or organization's purpose, its design, and the living system—the interaction of stakeholders—and the related outcomes. The difference between an organisation's planned outcomes and the actual outcomes is due to emergence. This difference should not be taken as unwelcome deviance, but as welcome information about the dynamics in the system, where it may be blocked, and where it may want to move. Generative Evaluation embraces emergence and makes it fruitful for reviewing and adapting the purpose, design, and the living system of a project, programme, or organization.

This paper shares insights from our experiences with Generative Evaluation. We are still experimenting with the approach, and so the intention of the paper is to invite readers to share their own experiences with similar approaches. We know that many evaluators are experimenting with similar methods to facilitate learning and change. We hope this paper can serve as a useful starting point to organize the various methods in the evaluation space and consolidate our experiences. The paper draws specifically on three projects as case studies, where we used a variety of generative approaches. Our conclusions are preliminary, yet we can clearly sense that there is high demand and interest in doing evaluation differently.

It may be no coincidence that we as authors are not full-time evaluators but rather facilitators of systems transformation. We have been supporting complex programming, mainly in the domains of inclusive markets, food systems, and natural resource management the past two decades, from design to implementation and evaluation and experimentation. As such, we feel the needs of the designer and the implementor when we evaluate, and we seek to create spaces for collective reflection, learning, and adaptation.

We have embarked on a journey since 2019 to design evaluation that supports living systems to meet their complexities. This paper draws on three case studies where we experimented with generative methods in learning and evaluation:

1. An *intergovernmental donor organization* that engaged us to assess and improve the evaluation systems of its 18 agencies using generative and systemic approaches.
2. Evaluating and refining a Theory of Change aimed at behavioural shifts of farmers in Brazil for a *communications agency*.
3. Evaluation and learning for a *global partnership* aimed at systemic change in global supply chains.

Whilst we are still experimenting, we have learnt that Generative Evaluation helps to:

- *Express and co-construct systemic issues*: Unexpressed and often unintended dynamics are allowed to surface.
- *Create connection and energy toward purpose*: The process is reinvigorating for participants, who connect deeply with others with a shared purpose and often emerge with fresh momentum.
- *Foster generativity by opening a world of possibilities*: Generative Evaluation can help learn from living systems and emergence to re-shape our purpose and how we achieve it.

In this paper we describe Generative Evaluation and its key properties and practices, and then draw on our three case studies to highlight the learning from our early-stage exploration of this approach.

What is Generative Evaluation?

Evaluation has become prerequisite of most publicly funded projects and initiatives and is used in many contexts to understand "what is working and what is not." Evaluation is widely regarded as the key mechanism for enhancing accountability and informing program development and planning through reflective observation, formative processes, and summative assessments of social change (Stockmann & Meyer, 2016). Typically, the evaluation is conducted by a third party, with the idea of engaging a neutral and objective observer. In this approach, learning is outsourced to the evaluator and, as a result, learnings and recommendations are often not easily absorbed by the stakeholders implementing the project or initiative. They simply have not been invited to follow along the process and are often stuck in routines that feel necessary.

This prevailing belief in the “gold standard” of evaluation—emphasizing neutrality, objectiveness, and transparency—has been central to evaluation frameworks since the 1990s (OECD, 1990), shaping how success is defined and measured, which in turn influences project design and implementation (Mayne, 2008). In fact, any approaches, principles and measurement logic of evaluation require a specific design and lock the living system in a particular purpose. In this way, evaluation affects organizational operations, values, learning processes,

and the extent to which experimentation, innovation, and honesty are encouraged. In essence, evaluation drives organizational culture, shaping what is valued, the dominant leadership styles, symbols, procedures, routines, and definitions of success (Cameron & Quinn, 1999).

The notion that policy choices will become easier and programs more efficient and impactful due to a norm of neutral, objective evaluation has come under increasing scrutiny. Rather than rely on expert judgment, there is widespread understanding that participants of social change need to learn collectively and be bolder in their approaches to meet and experiment with the complexities and ambiguities of living in a globalised world (Patton, 2013; Senge, 1991). Further, in our era of rapidly changing communication, marked by fake news, conspiracy theories, and emerging learning technologies, citizens are questioning expert information and notions of neutrality or objective truth underpinning policy response (Hense, 2018, p. 3) as was the case with citizen responses during the COVID-19 pandemic.

The future of evaluation will require more than the expert evaluator or scientist as the sole arbiter of evidence and truth. Policy, programmatic, or citizen approaches aspiring to change systems through complex interventions involving a multitude of actors may need to ask themselves: how can these interventions learn both with and for living systems? Indeed, isn't systemic change fundamentally about how living systems are learning and responding collectively?

We have been engaged in generative practices in evaluation and learning, which have the potential to foster a more dynamic and supportive organizational culture. Generativity in our practices is the ability to produce or create something. In psychology, Erik Erikson first coined the term in the 1950s to describe the developmental stage where adults begin to guide the next generation (Erikson, 1978). American social psychologist Kenneth Gergen coined the concept of generative theory as a scientific approach to challenge common assumptions and open up new possibilities and forms of action (Gergen, 1979). Gergen argues for an approach to theorizing that is free from the immediate need for verification and facts and can thus "generate" desirable social outcomes. He links his proposal with the rationalist tradition going back, among others, to Kant and Hegel, who regard reason, rather than empirical facts, as the source of knowledge.

Peter Senge builds on this tradition in his seminal 1990 article (Senge, 1990). Generative learning, he writes, enhances our capacity to create. It requires new ways of looking at the world, seeing the systems that control events, and also seeing clearly where we want to be. Our "vision," and telling the truth about where we are in our "current reality," helps to foster a learning organization. Otto Scharmer, co-founder of the Presencing Institute, lead author of the Theory U approach, and a close collaborator of Senge, picks up these ideas in his concept of "generative dialogue" (2002). Generative dialogue transcends individual conversations to become a collective experience of meaning-making,

discovery, and transformation. Meaning flows freely, and a generative flow unfolds, unleashing collective creativity through authentic and evolving identities. It's a space where conversations cease to be routine and become a force for genuine, transformative change. Generative Evaluation builds on this tradition and leverages intuitive, systemic, and actor-based methods from various disciplines and practices, including psychology (Perls et al., 1951), ancient wisdom (e.g., Macy & Brown, 2014, Yunkaporta, 2019), coaching, Theory U (Scharmer, 2016, 2018), Design Thinking (e.g., Brown, 2008), and Systems Practice (e.g., Senge, 2014) (see Table 1). We integrate these generative practices into our evaluation work to foster more dynamic and transformative processes.

Our basic assumption in Generative Evaluation is evaluation as dynamic and transformative processes and, with this approach in mind, deviations from the purpose and design of a program provide useful indications of important factors in the living system (Figure 1). Traditional evaluation methods tend to see these deviations as failure that needs to be corrected. By contrast, a generative approach embraces emergence. Evaluation is utilized as an opportunity to explore the interactions between purpose, design, and the living system and to update and reinvigorate all elements, building on the work of Michael Patton (Patton, 1994). With Developmental Evaluation, Michael Patton has indeed introduced the concept of emergence into the evaluation of complex problems (Patton, 1994). This approach promotes organizational development and innovation through evaluation and learning, emphasizing the importance of facilitated processes (Patton, 2013). Generative evaluation builds on this lineage and enriches it with generative methods emanating from the work of Peter Senge (1990, 2014), Otto Scharmer (2016, 2018), and others. In other words, while Developmental Evaluation primarily relies on an analytical approach using qualitative and quantitative data, Generative Evaluation broadens this by incorporating intuitive, embodied, and co-creative practices to observe and move with emergence.

While still in development, some aspects of Generative Evaluation have become clear from testing approaches in our case studies. Generative Evaluation explores the possibilities of what is emerging or wants to emerge (Figure 1). Instead of solely focusing on explaining the past, it allows participants to investigate the interdependencies and interconnectedness within emergent processes. Generative Evaluation explores emergence in a living system, say a project, initiative, or organization, by observing the relation between the purpose, design, and living system (Figure 1). All three dimensions can be adjusted in the process to integrate learnings from the past and move towards a desired future.

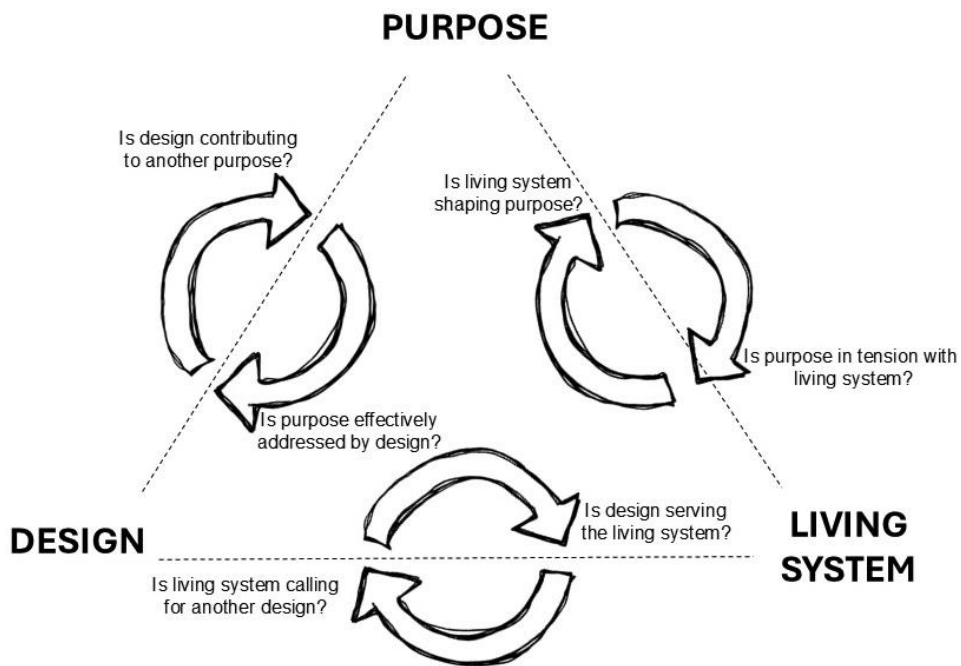


Figure 1: Interactions within emergent processes between purpose, design and living system.

Generative Evaluation explores and updates the interaction between purpose, design and the living system. The *living system* refers to the dynamic interconnections and interdependencies between the various people and contexts that are involved in or affected by the purpose and the design of the project. Values, power structures, needs, incentives, and constraints reflect in these relationships. The *purpose* is the explicit and implicit intention and function of the project or organization to be evaluated. In addition to any formalized purpose, we may observe different actors have at times different views of what the purpose of the project or organization is. The *design* of the project or organization describes the structures and processes to achieve the purpose and manifest it in the living system. This includes the Theory of Change, organizational setup, workplan and budget, procedures and so forth. Generative Evaluation helps us learn about how well the design serves the intended purpose in real life; to what extent the purpose is actually shared by various actors; and what the living system teaches us about adjusting the purpose and design.

Principles and Practices of Generative Evaluation

Generative Evaluation principles are co-creative, non-judgmental, intuitive, and systemic (Figure 2). The principle of *co-creative* helps to bring together diverse perspectives, encourages shared ownership, and fosters creativity and innovation, which all together help participants to keep an open mind and be *non-judgmental*. The principle of *intuitive* help builds perspectives, evidence, data, and interpretation not only from the lens of conscious reasoning, but also from feelings and felt sense. The principle of *systemic* looks at interdependencies and the whole. Some practices we used are more *systemic* while others are more

actor-centered, looking at a particular actor in the system. It is through a set of generative practices drawing from disciplines and practices, including psychology, ancient wisdom, coaching, design thinking, systems practice, facilitation and organisation development, that those principles are realised (Figure 2; Table 1).

Definition of those four principles in Generative Evaluation:

1. *Co-creative* practices enable participants to create outputs together. In addition to providing own input, the task is to listen to the input of others and to collectively develop an idea, a hypothesis, a proposal, an experiment, assessment, or other output.
2. *Non-judgmental* practices encourage participants to keep an open mind about the purpose, design, and living system features. In a constructive spirit, assessing what is “emerging” and what “works” is a shared process, rather than a benchmark against an objective target or ideal.
3. *Systemic* practices encourage participants to focus on interdependencies rather than on individual elements of a project, program, or wider system. These practices also enable participants to integrate issues or topics that are outside the defined scope of the project. This wider perspective helps to contextualize findings and draw conclusions about why things are working or are not working, including with relationships, values, power structures, needs, incentives, and constraints.
4. *Intuitive*: In addition to cognition, practices leverage the intuition of participants. Tapping into emotions, images, our sense of space, music, and poetry, as well as fantasy, enables participants to share insights that integrate their experience and insights that draw from processes other than conscious reasoning, regarding the complex living system and how it relates to the design and purpose.

We characterize each practice from Figure 1 briefly here in Table 1. A full description would go beyond the scope of this paper, but many useful resources on these practices are provided in Table 1.

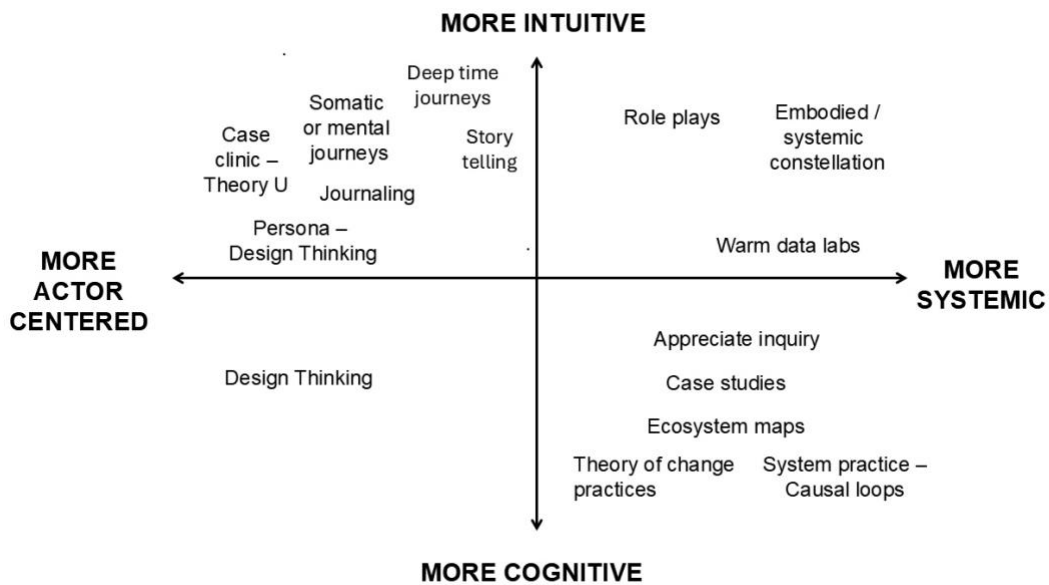


Figure 2: Generative Practices in Generative Evaluation

Practice	Description	Reference
Appreciative inquiry	Stimulates collective inquiry into the best of what is, in order to imagine what could be, followed by collective design of a desired future state that is compelling and thus does not require the use of incentives, coercion, or persuasion for planned change to occur.	Busche, 2013 Whitney & Cooperrider, 2011.
Design thinking	A process for creative problem solving. Methods put the user of a solution into the center. They also include many creative exercises that enable participants to brainstorm together and quickly produce tangible outputs based on real-life feedback.	Brown, 2008
Deep time journeys	Involves the process of experiencing our own personal lives as held within a larger context of space and time, widening our perspective to include the wisdom of our ancestors and our children’s children. As a result, we open ourselves to a deepened appreciation of the evolutionary journey we have been through as a species. It is inspired by Indigenous practices such as the seventh-generation principle and practice which is based on an ancient philosophy of the Haudenosaunee Confederacy (Iroquois), or Aboriginal dreaming and storytelling.	Haudenosaunee Confederacy (n.d) Macy & Brown, 2014 Yunkaporta, 2019
Ecosystem maps	Portrays the actor landscape relevant for a specific project or question and the interrelations between these actors. They can also highlight barriers and incentives that work on these actors.	
Embodied/systemic constellations	Uses our intuitive sense of space to map interdependencies between actors or topics. As representatives position themselves in the room, tensions, imbalances, alliances, and other	Whittington, 2012 Arts et al., 2021

E.g. 3 and 4 D mapping	“constellations” become apparent. Constellations can also be used to “prototype” changes and quickly see how they might affect other actors in the ecosystem. It is inspired from family constellation and embodiment practices.	
Journaling	The practice of writing down your thoughts, feelings, insights, and more. Used as a generative workshop practice, the facilitator guides the participants through a series of questions. Participants write down their reflections onto their note pad, just for themselves. Questions connect participants deeper with their emotions and insights. A round of reflection in the group then enables others to benefit from the insights.	
Somatic/mental journeys	Used to connect participants with themselves and their emotions and ground them in the present moment. Mental journeys are also used to explore certain aspects intuitively, such as perspectives of relevant stakeholders, or future developments of a project. It is inspired from mindfulness and Gestalt therapy.	Perls et al., 1951
Story telling with metaphors and symbols	Used to synthesize insights into an intuitively accessible and meaningful “shortcut.” For example, an evaluation of a project aimed at a regenerative economy may use the symbol of a seed to explore the theory of change of generativity during the evaluation.	
Persona	Holistic portraits or archetypes of fictional characters that represent customers, users, or other interest groups. Personas are used to personify the needs of the interest groups. This helps gain empathy and a better understanding of the users. Persona is a key aspect of Design Thinking.	Brown, 2008
Role play	Enables participants to explore different stakeholder perspectives as well as interdependencies between actors with empathy. Participants take on different roles and then interact around a certain subject or issue. Role play can also be used to quickly prototype solutions and explore how they might be perceived by different stakeholders. Role-playing helps with creativity, self-awareness, group cohesion and “out-of-the-box” thinking.	Bowman, 2010
System practice	A methodology to identify and map different forces related to an issue, subject or question, and the causal links between them. The output is a causal loop map that connects factors into self-generating cycles. Maps can be used to identify leverage points as areas for intervention.	Senge, 2014
Case Clinic of Theory U	An approach where a case-giver presents a case, and peers or team members act as coaches tapping into not only cognitive reasoning but generative and felt senses.	
Case studies	A detailed study of a specific subject that illustrates a certain point or enables participants to engage with a specific question or opportunity.	

Theory of Change	An approach that consists of a variety of ways of developing a causal model that links the program's activities to a chain of intended or observed outcomes, and then using this model to guide the evaluation. Participants can engage with the Theory of Change, explore assumptions, test causal connections, and adapt it.	Stein & Valters, 2012
Theory U	Theory U is a methodology that enables organizations and other groups of stakeholders to collectively move towards a desired future in five steps, or movements: co-initiating, co-sensing, presencing, co-creating and co-evolving. Processes using this approach are also often called "generative" processes. The U movement is the basis for all our facilitation design.	Scharmer, 2016
Warm Data Labs	Group processes in which information and formulation of cross-contextual knowing is generated—the "Warm Data." For example, to understand a family, one must understand not only the family members, but also the relationships between them, and the economic, education, media, health, and other systems they are embedded in.	Bateson, 2016

Table 1: Summaries of Generative Evaluation Practices Used.

Case Studies

We illustrate the principles and practices of Generative Evaluation by drawing on three case examples from our work including the evaluation task, what approach we took, and what we learned.

Case 1: Assessing Evaluation Systems

Evaluation Task

A large intergovernmental donor organization engaged us to assess and improve the evaluation systems of 18 international development agencies receiving their funds.

Our Approach

Our proposed methodology incorporated systems thinking to map the evaluation systems of different agencies, focusing on their learning and accountability processes. This mapping was informed by detailed case studies, which would identify learning opportunities for applying design thinking. We also proposed using design thinking to understand the agencies' perspectives and needs. This approach aimed to uncover systemic issues, identify opportunities for improvement, and co-create solutions. Additionally, design thinking included the co-design of solutions with the agencies to enhance their evaluation systems.

In our first co-design workshop, we invited evaluation managers to journal their pain points and the needs of their evaluation systems. This was a novel experience for staff who were typically focused on technical questions, evaluative opinions, and judgments within the context of evaluation. The exercise fostered a sense of community as similar pain points emerged across organizations. It became evident that everyone in the room deeply cared about evaluation, and the connection through this shared purpose was palpable.

Our generative practices throughout the co-design process included journaling exercises, with classic prompts such as “What excites you?” and “What frustrates you about your evaluation system?”

What We Learned

This approach allowed the agencies to be authentic and openly discuss barriers within the current compliance-based system. One significant issue flagged was that the system creates almost no direct incentives for candor with the donor or for deeper learning. Agencies questioned why they would be candid and foster learning when doing so could jeopardize future funding. Consequently, they revealed how difficulties and failures in projects might be concealed to protect future project opportunities, rather than using these experiences as valuable learning opportunities. By acknowledging such systemic issues—the elephant in the room—in their evaluation systems, agency staff were empowered to make their own recommendations during the design thinking sessions.

In addition to using design thinking and generative practices, we were also mapping the evaluation systems of the agencies. Our goal was to explore the feedback loop between the donor's evaluation system and that of their beneficiaries. However, from the outset of the contract, the donor made it clear that our mandate was to assess the agencies' evaluation systems, not the donor's own system in relation to its beneficiaries. Despite this, during the generative practices and design thinking sessions, agencies naturally revealed the real-life conditions and constraints of their evaluation systems, many of which were closely tied to the donor's rules and incentives. Our approach generated honest conversations on barriers to learning due to asymmetric power relationships between donor and agency, and some solutions to overcome this.

Case 2: Making Theory of Change Personal

Evaluation Task

In another case study from Brazil, we were assisting a communications agency in refining their Theory of Change (ToC) in preparation for a multi-year intervention following a pilot phase. The initiative's long-term goal was to reduce deforestation by farmers and agribusinesses in the region, with a focus on using communication to encourage behavioural change among farmers. Rather than

evaluating the pilot phase itself, our task was to assess the current Theory of Change using data gathered during that phase.

Our Approach

By fostering empathy and understanding for the target audience, we aimed to create a more responsive and effective approach to inviting behavioural change

As part of our methodology, we utilized data collected on the target audience during the pilot phase to develop personas—artefactual characters synthesized from the target groups' key attributes such as values, needs, pains, and longings. Personas help us imagine and empathize with the people targeted or affected by the change we aim to make. These personas were then employed to refine the Theory of Change, making assumptions about behavioural change more explicit. Additionally, we used role play allowing our client to take on and explore perspectives of the target group. This helped them to sense their own position in the system which had triggered resistance by farmers in the past. Empathizing with this resistance and connecting with values of the target group, our client learned that they would need to show up and communicate differently from the stereotypical environmental activists they had been seen as.

What We Learned

As a result of the generative approach, the client team achieved a new understanding of their target audience and the conditions necessary for change. Eventually, the communication agency jointly reshaped the Theory of Change to become more responsive to the living system they set out to engage with. This was particularly valuable in the polarized context of conservation and deforestation in the Amazon, where there was a risk of reinforcing existing assumptions and biases, potentially deepening the divide between farming and forest protection.

Instead of employing a one-way behaviour change model, the client acknowledged the agency of the target group. They could appreciate farmers, their views, values, and needs in a different way, thus informing their future approach of communicating for change.

Case 3: Learning for Systems Change

Evaluation Task

Our third case study is a learning and evaluation project for a global partnership aimed at systemic change in global supply chains. We served as evaluators for a midterm evaluation, and later as facilitators for systems change. This global initiative involved substantial investments aimed at reducing deforestation within major agricultural supply chains. The focus was on creating enabling

conditions for sustainable production, which included providing financial incentives and fostering demand for sustainable products. The partnership brought together some of the world's leading NGOs and UN organizations dedicated to forest conservation and sustainable agriculture.

Our Approach

This intervention started with a traditional mid-term evaluation of a large multi-country effort to create systemic change in global supply chains towards reduced deforestation led by multiple international development agencies. We expanded our intervention into system practices such as system mapping and embodied constellation of specific supply chain and their challenges and learning workshops, where we applied practices, such as appreciative inquiry and storytelling with metaphors and symbols, aligned to their Theory of Change, and prototyping through role plays of next possible design. We held systems practice workshops in the two intervention countries—Brazil and Paraguay—to map the system and identify leverage points. We held a global learning workshop at the end of the programme.



Figure 3: Use of Metaphors and Symbols in Generative Evaluation to Stimulate Storytelling.

In this case, the symbolic graphic depicted the global partnership's theories of transformation, which in our learning workshop, helped stimulate storytelling around the challenges of this Theory of Change.

What We Learned

The partnership struggled with its overall theory of transformation for integration of demand, supply, and production. We captured this struggle in the image we created for their learning workshop, which was a graphic of three hands (supply, demand and production) which strived to connect, but could not entirely be integrated (see Figure 3). This symbolic graphic evoked strong emotions during the learning workshop, even among staff from the donor agency who had been instrumental in designing the partnership and its theory of change. As a result, intuitive insights including pain points around the Theory of Change were revealed. For example, the donor's insistence on fostering collaboration between traditionally competing agencies through this project which aimed to promote the idea of integration, inadvertently constrained all parties involved. Despite these challenges, the partnership itself was still appreciated and welcomed.

The systems practice and learning workshops also revealed that the partners were overwhelmed by the complexity of global supply chains and their political, economic, social contexts. We were struck by repeated feedback from participants, including those with over 30 years of experience in development banks, NGOs, and similar organizations, that the “context for systemic change in global chain is too complex.”

Nonetheless, systems practice gave a big picture of where each partner in the partnership was needed for systemic change in global supply chains (from demand, to supply, and finance) to work. It showed that different theories of change were needed as conditions for systemic change in global supply chains were challenging. This reduced competition among partners on the “right” approach. Embodied constellations and role plays helped the partnership understand the rationality of the stakeholders, their “why” of behaviours, and the conditions that lead to systems behaviours. This helped all partners improve the design of the partnership.

The biggest pain point of the partnership was the lack of flexibility and capacity to foster social innovation and experimentation with business actors within the existing design, given the complexity of supply chains. This was due to the rigid result framework of non-changeable targets and outcomes, which did not encourage experimentation, innovation, and adaptation. During the learning workshop at the end of the partnership, we organized a role play to test their ideas of more generative design—specifically, programs and partnerships that support greater experimentation, innovation, and adaptation. Participants highlighted design principles around being adaptive and iterative, about nurturing and facilitating collaboration and behaviour change processes and around co-creative design process for outcomes and impacts. For the design principles to be realized, participants identified the need for skills and incentives that support collaboration, facilitation, systems thinking, social innovation, and new approaches to monitoring, learning, and evaluation for systemic change. During our role plays, designed to simulate how the design principles would be

received, many participants playing the roles of donors and government representatives reacted negatively. They expressed concerns such as, "We're already locked into themes and indicators for the next funding cycle," or "We need to stay focused on results and impacts," or "Supply chain stakeholders and governments want practical solutions, not abstract design principles." Despite this resistance, seeds of possibility were planted while everyone still faced the realism of the hard conditions for change.

In this way, generative practices enabled the partnership to better prepare for and work more realistically with future initiatives. Generative practices highlighted the need for additional resources, enhanced skill sets, and incentives for collaboration, system approaches, and social innovation and experimentation.

Early Findings

We have only started to bring generative practices into evaluation processes. From our early experience, the results it yielded, the feedback we received from clients and participants, and the energy in our workshops, shows that these practices bring a new and highly welcome twist into evaluation. Most organizations now want to use evaluation for learning, in addition to accountability (Meyer & Stockmann, 2016). Often, this learning is conceived as an analytical exercise, where an evaluator draws conclusions and feeds them back as recommendations. We promote a more collective approach to learning, a journey from the past experience to the future design. We see in these and other examples how such an invitation to "generate" and to "re-generate" is changing the dynamic at three levels. It allows tensions to surface, invites connection in the group, and creates space for future opportunities.

Express Systemic Issues

In the cases above, generative methods have helped surface systemic issues, which, while they may seem obvious, are often obscured by unconscious "organizational silence" (Morrison & Milliken, 2000). In the case "Assessing Evaluation Systems" for a donor (Case Study #1), it was quite obvious to all participants that candor was disincentivized by donor-agency power asymmetry and hence effective learning was obstructed. The process allowed this "elephant in the room" to be expressed, which allowed the group to look for solutions. Likewise, in the case "Learning for Systems Change" (Case Study #3), it was plain that a rigid result framework of non-changeable targets and outcomes was incompatible and non-conducive for experimentation, innovation, and adaptation. Acknowledging this pressure point enabled the team to look at their achievements with a new perspective and appreciate the learnings more. Finally, in the second case "Making Theory of Change Personal" (Case Study #2), the communications agency needed to realize that they were also part of the system, and from their target group's perspective, that they were even part of the problem. Becoming aware of this, they moved on toward acknowledging the

realities and agency of their target group. This allowed them to shift how and from which vantage point they would be advocating for change from then on.

Silence about tensions persists due to unquestioned belief systems, power structures, incentives, and a lack of psychological safety (Edmondson, 2018). By enabling participants to co-construct these systemic issues in a safe enough space, they could surface them. In this way, these issues became seeds for designing meaningful co-improvement. By creating a safe enough space for participants to co-construct these systemic issues, they were able to bring them to light. In this way, these issues became seeds for meaningful co-improvement. For instance, in the first case study, agencies recommended cross-agency learning processes to encourage open dialogue and shared learning, ensuring this knowledge was accessible to all stakeholders. Such cross-agency collaboration can help overcome systemic barriers to learning, particularly when fear from power imbalances between donors and beneficiaries impede candor and learning (Case Study #1).

Create Connection and Energy for Moving Towards Purpose

Evaluation is often associated with judgment and criticism. But what if we introduced more appreciation into the process? In our work with multiple agencies and their evaluation systems (Case Study #1), we described and categorized the different evaluation systems and their best practices, rather than judging “what is wrong” or comparing “what is better.” This appreciative approach encouraged evaluation managers from different agencies to engage with one another’s evaluation systems in the spirit of connection and learning. We heard participants make comments such as “We don’t use theory of change as part of our evaluation approach, but we want to learn how you have integrated evaluation of Theory of Change in your evaluation”; or “We conduct in-home terminal evaluation which brings quality but that leaves us no time for learning, how are you doing it differently”; or “Our rating systems are different, so how can we have similar ones to compare across similar portfolios” (various participants, Case Study #1). This appreciative approach fostered a willingness to learn from each other, identified common needs, challenges, and objectives, and led to brainstorming potential opportunities in their evaluation systems. These interactions also created connective tissue between the different parts of the living system, fostering a sense of shared learning and a collective journey.

We are finding that when a living system is stuck, Generative Evaluation can help actors experience each other, their work and their organization, in new ways. It revitalizes and strengthens the relationships within the system, reconnecting them with the purpose and design of the initiative. Because generative methods are highly co-creative and often intuitive, they provide stakeholders with opportunities to interact openly and share their perspectives. These intuitive practices allow participants to explore and express their own emotions, life experiences, insights, and dreams, making these personal reflections a key part of the evaluation process. For example, playful, intuitive,

and empathetic approaches with use of persona and role-play helped the communication agency reflect on their assumptions and biases towards farmers in their theory of change (Case Study #2). Owning these assumptions and biases made it possible for the communications agency to re-connect with farmers and adapt their communication and behavioural change strategy to a more relational one. In other words, the focus shifted from merely analysing data and insights to integrating each participant as a whole person including their experiences, mental models, and feelings into the Theory of Change and that helped to re-connect to stakeholders and move towards new strategy and purpose.

We also observe that clients and partners often lack a safe space and the opportunity to acknowledge and express the current reality, including pain points and collective dreams. This absence of space for recognizing "what is" contributes to a state of stagnation, as there is often excessive focus on "what should be." For example, in the case "Learning for Systems Change," we created a symbolic graphic image (Figure 3) to depict the partnership's struggle with their theory of change (Case Study #3). By incorporating symbolism and emotions into the evaluation and learning process, it allowed both donors and implementers to tune in on "what is"—that is the current reality—rather than being caught up in the judgmental aspects of "what should be." Traditional evaluation frequently reinforces this focus on "what should be," which is why it often fails to help organizations, teams, or projects overcome their challenges and get unstuck. Expressing the "what is" brings a sense of lightness, enabling more meaningful connections. This, in turn, fosters greater energy, enthusiasm, motivation, clarity, acceptance, and forgiveness for failures and constraints. Such meaningful connections enhance the potential for learning and pave the way for collaborative efforts and co-creating solutions. As a result, the living system becomes revitalized and embarks on a dynamic learning process, actively searching for its future potential.

Fostering Generativity in a World of Possibilities

In the *Art of Possibilities*, the Zanders write that "in the measurement world, you set a goal and strive for it" (Zanders & Zanders, 2002, p. 21). In other words, success or impact is viewed only from the point of view of that goal. In the universe of possibility, the Zanders continued, "you set the context and let life unfold" (Zanders & Zanders, 2002, p. 21). In other words, if we want to learn about how our intended strategies are playing out in complex systems, we would be wise to observe emergence in the context that we set.

In the case "Learning for Systems Change," emergence and adaptation to deal with complexity in transforming global supply chains was much needed but not possible because of rigid results framework (Case Study #3). In reflecting on this, participants including implementation agencies and donors identified design principles they wanted to use for the design of the next program and funding cycle to foster systems change. When prototyping those design principles using role-plays, it became clear that the system was not ready to foster systems

change. We heard role playing participants, especially those embodying roles of donors and government representatives, making comments such as, "We're already locked into themes and indicators for the next funding cycle," "We need to stay focused on results and impacts," and "Supply chain stakeholders and governments want practical solutions, not abstract design principles." Generative Evaluation helped participants look at nourishing conditions for "ready-ness for change" (Bateson, 2022) as seeds of possibility while still facing the realism of the hard conditions for change. Nora Bateson calls this a theory of change that changes a theory of change (Bateson, 2022).

By observing emergent feedback and resonance of the living system, Generative Evaluation has the potential to connect evaluation with strategy. Honouring these responses and the agency of its actors allows us to co-create and update the purpose and design that can function more adequately with the living system—not against it. This is what we mean when we say evaluation needs to be a resource for the living system. Generative Evaluation invites the living system as the main source of generativity to help us re-define success and impact in line with its living complexity.

Conclusion

Generative Evaluation introduces tools and processes for stakeholders to explore how and why a living system targeted by a project or initiative is actually responding to its purpose and design. Inviting these learnings into a generative process allows people and organizations to jointly question and co-create why, what, and how a certain change may be brought about in the future.

In fact, conversations and insights in Generative Evaluation often revolve around the limits related to current design, including structures and processes, aiming to bring about change. Moreover, previously unvoiced, but highly influential aspects of a living system such as asymmetric power relationships become more tangible and demand their way into our why, what, and how we do things. These results are challenging. Changing mindset and culture does not happen with one evaluation. But if more people start to use generative practices and allow for these deeper systemic issues to be spoken during evaluation, the more we foster a culture of openness and learning among important institutions. While change is often met with resistance, generative practices create a safe space where change can be explored as a group.

We are only at the beginning of testing and exploring generative practices in evaluation processes, which can take place throughout the evolution of a project, partnership, community, or organisation. We already see that it brings value to the process at different levels and helps to synchronize the purpose and design while learning from and with the living system they engage with. It leverages the established routine of evaluation to learn together, articulate and dissolve tensions, strengthen relationships, and generate possibilities for the future.

Evaluation systems are what ultimately define how projects and organisations look. “What gets measured gets done” is a widely known management proverb. The indicators and measurement logic require a specific design and lock us in a particular purpose. Changing evaluation therefore means changing how projects and organisations look, but also what “success” or “impact” can look like. If we are truly aiming for lasting impact, we must become better at designing evaluation for living systems to meet their complexities rather than projecting our strategies onto them in a one-way street. If we do that, evaluation and strategy will not be separate processes. Generative Evaluation provides a concrete and practical way to move a step closer in that direction.

References

- Arts, J., Baldini, A., Goodman, M., Hayashi, A., Jandernoa, B., & Scharmer, O. (2021). Exploring the origins of practice: In dialogue with founding faculty of the Presencing Institute. *Journal of Awareness-Based Systems Change*, 1(1), 125–137. <https://doi.org/10.47061/jabsc.v1i1.675>
- Bateson, N. (2022). An essay on ready-ing: Tending the prelude to change. *Systems Research and Behavioral Science*, 39(5), 990–1004. <https://doi.org/10.1002/sres.2896>
- Bateson, N. (2016). *Small arcs of larger circles*. Triarchy Press.
- Bowman, S. L. (2010). *The functions of role-playing games: How participants create community, solve problems and explore identity*. McFarland.
- Brown, T. (2008, June). *Design Thinking*. Harvard Business Review. <https://hbr.org/2008/06/design-thinking>
- Bushe, G. R. (2013). Generative process, generative outcome: The transformational potential of appreciative inquiry. In S. Evans (Ed.), *Organizational generativity: The appreciative inquiry summit and a scholarship of transformation* (Vol. 4, pp. 89–113). Emerald Group Publishing.
- Whitney, D., & Cooperrider, D. (2011). *Appreciative inquiry: A positive revolution in change*. Berrett-Koehler Publishers.
- Cameron, K. S., & Quinn, R. E. (1999). *Diagnosing and changing organizational culture: Based on the competing values framework*. Jossey-Bass.
- Edmondson, A. C. (2018). *How fearless organizations succeed*. Strategy and Business. <https://www.strategy-business.com/article/How-Fearless-Organizations-Succeed>
- Erikson, E. H. (1979). *Dimensions of a new identity*. W. W. Norton & Company.
- Gergen, K. (1978). Toward generative theory. *Journal of Personality and Social Psychology*, 36(11), 1344–1360. <https://psycnet.apa.org/doi/10.1037/0022-3514.36.11.1344>
- Haudenosaunee Confederacy. (n.d.). *Haudenosaunee Confederacy: Values*. <https://www.haudenosauneeconfederacy.com/values/>
- Hense, J. (2018). German angst regarding the future of evaluation. In O. W. Andersen, & K. Williams (Eds.), *Evaluation Connections* (pp. 2–3). European Evaluation Society.
- Macy, J., & Brown, M. Y. (2014). *Coming back to life: The guide to the work that reconnects*. New Society Publishers.

- Mayne, J. (2008). Fostering an evaluative culture: The key to effective evaluation and results management. *Canadian Journal of Program Evaluation*, 24(2), 1–30.
- Meyer, W., & Stockmann, R. (2016). *The future of evaluation: Global trends, new challenges, shared perspectives*. Springer.
- Mintzberg, H., & Waters, J. A. (1985). Of strategies, deliberate and emergent. *Strategic Management Journal*, 6(3), 257–272. <https://www.jstor.org/stable/2486186>
- Morrison, E. W., & Milliken, F. J. (2000). Organizational silence: A barrier to change and development in a pluralistic world. *Academy of Management Review*, 25(4), 706–725. <https://doi.org/10.2307/259200>
- OECD. (2010). *Quality standards for development evaluation*, DAC guidelines and reference series. OECD Publishing. <https://doi.org/10.1787/9789264083905-en>
- Patton, M. Q. (1994). Developmental evaluation. *Evaluation practice*, 15(3), 311–319. <https://doi.org/10.1177/109821409401500312>
- Patton, M. Q. (2013). The future of evaluation in society: Top ten trends plus one. In S. Donaldson (Ed.), *The future of evaluation in society: A tribute to Michael Scriven* (pp. 45–62). Information Age Publishing.
- Perls, F. S., Hefferline, R., & Goodman, P. (1951). *Gestalt therapy: Excitement and growth in the human personality*. The Gestalt Journal Press.
- Scharmer, C. O., Arthur, W. B., Day, J., Jaworski, J., Jung, M., Nonaka, I., & Senge, P.M. (2002). *Illuminating the blind spot: Leadership in the context of emerging worlds*. Society for Organizational Learning.
- Scharmer, C. O. (2016). *Theory U: Learning from the future as it emerges*. Berrett-Koehler Publishers.
- Scharmer, C. O. (2018). *The essentials of Theory U: Core principles and applications*. Berrett-Koehler Publishers.
- Senge, P. M. (1990). *The art and practice of the learning organization*. Broadway Business.
- Senge, P. M. (2014). *The fifth discipline fieldbook: Strategies and tools for building a learning organization*. Crown Currency.
- Stein, D., & C. Valters (2012). *Understanding theory of change in international development*. The Justice and Security Research Program.
- Stockmann, R., & Meyer, W. (2016). The future of evaluation: Global trends, new challenges and shared perspectives. In R. Stockmann, & W. Meyer (Eds.), *The future of evaluation: Global trends, new challenges, shared perspectives* (pp. 9–19). Springer.
- Whittington, J. (2012). *Systemic coaching and constellations: An introduction to the principles, practices and application*. Kogan Page Publishers.
- Zander, R. S., & Zander, B. (2002). *The art of possibility: Transforming professional and personal life*. Penguin.
- Yunkaporta, T. (2019). *Sand talk: How Indigenous thinking can save the world*. Text Publishing.