SOCIAL FINANCE

Changing Lives,
Changing Systems

Navigating system change evaluation

A white paper

socialfinance.org.uk

Contents

Introduction **3**

Why is evaluating system change so hard? **4**

Systems don't change fast 4

System outcomes are often emergent 5

Causal relationships are 'foggy' in complex systems 5

The nature of the intervention is dynamic 5

There are multiple views on what constitutes both the system and success 6

Evaluation is dead! Long live evaluation **7**

Four ways evaluation needs to look and feel different in system change contexts 7

Not throwing the baby out with the bathwater – three ways evaluation will stay the same 10

System change evaluation in practice **16**

Embed evaluative thinking into your strategic planning 16

Plan for iteration, but don't lose your 'North Star' 17

Check for impact regularly, but put it in (system) context 18

Retain rigorous, causal thinking but be sensitive to emergent outcomes 19

Treat your strategy as a hypothesis and learn with direction 20

How to know you're making progress? An inquiry framework for system change **22**

Good evaluation as the 'engine' of system change **28**

Acknowledgements **29**

Interviewees and collaborators **30**

Further reading **31**

Authors **32**

Endnotes **33**

Introduction

At Social Finance, we are interested in achieving 'big outcomes' – healthy populations, a fairer world, strong economies, a more sustainable planet. We know that you can't get there without diving 'under the hood' of the systems that generate these outcomes to tinker with the wiring, change the fuel, or even completely reinvent the way they work and the purpose they serve.

A lot of what we have learned over the last decade about what those big system shifts look like in practice was captured in our report Building Routes to Scale. But as we have got clearer on what we think it takes to make change at scale, it has also become more urgent to ask ourselves how we can know whether we are changing the system in the first place, and whether the changes we are making are having the impact we desire.

We are not alone in feeling that evaluating system change work is notoriously hard. Last year, we spent six months in partnership with the System Innovation Initiative (SII), supported by the Rockwool Foundation, exploring how different players in the field of social innovation are grappling with these questions, and what the implications might be for funders, governments and 'system changers' who want to evaluate their work, or the work they are funding. Like us, other practitioners and their funders wonder how to marry what they have learnt about the nature of system change – its complex causal webs, decade-long timelines, and sheer scope – with the approaches and outputs of what we are accustomed to seeing as good, rigorous evaluation.

Across the impact sector, practitioners, funders and evaluators are wondering what to discard and what to keep as they reexamine evaluation practices in a new systems context. But while there are no 'holy grail' answers that meet all demands without compromise or tough decisions, the field of system change evaluation is far from barren. A growing group of evaluators and practitioners are involved in a generative conversation, continuing to evolve, refine and share approaches and developing a picture of how evaluative thinking can be an engine for real, lasting change. We have benefited hugely from engaging with them.

This report draws on that thinking, along with our own work in recent years on large-scale system change projects. It shares key lessons for how to approach evaluation in system change contexts – as well as our emerging thinking on how to apply this in practice.

Our focus for this work has been what could be useful to practitioners like ourselves, and to the organisations and funders we partner with – rather than seeking to contribute to the academic discussion. Our basis is our own experience of grappling practically with the challenges and tensions that inevitably emerge when theory hits reality, and our inclination is always towards pragmatism. We know that it can be hard to find space in tight budgets for formal evaluation activities, or to plan evaluations that meet the needs of every stakeholder. As a result, we don't pretend to present an 'ideal approach', but rather to propose ways of thinking that can be used by anyone working on system change to think more deeply about how to approach evaluation. We hope that this report will help would-be funders and commissioners to think more deeply about the value of their investment, help practitioners and funders to have more useful conversations with each other about impact reporting and accountability, and provide practitioners and evaluators with some framing and approaches to drive forward system change.

We cover:

- **Why is evaluating system change so hard?** We explore the key reasons why system change work appears to sit in tension with more familiar approaches to evaluation
- **How should evaluation adapt – what to lose and what to keep?** The changes that funders and practitioners of system change will need to make in their expectations of evaluation and how it works – as well as some core foundations that should remain intact, even if they look and feel a little different
- **Putting it into practice** What it might look like in practice for funders and evaluators to adapt their approaches – balancing the 'old' and the 'new' of evaluation – in the system change context?
- **Our emerging thinking – some potential approaches**: how you might assess whether you're making progress in shifting the system in meaningful ways.

Why is evaluating system change so hard?

Projects that have the ambition to 'change the system' tend to look and feel different. Compare a six-month mentoring programme intended to get more people into work in a given region, with an ambition to reshape the way the labour market works in the UK. With the former, we know the 'object' of evaluation (the mentoring service), the time-frame (six months) and the domain of impact (service beneficiaries in the local area in which the service operates). In system change approaches like the latter, the 'effort' to change a system seeks to reach beneficiaries directly and/or indirectly, takes place over a longer or as-yet undefined timeframe, seeks impact on a whole system, and can take many shapes – a network of aligned actors, a catalytic pilot venture, a portfolio of connected interventions.

In our experience of working to change systems, evaluating such efforts can be challenging, not necessarily because we lack the right methods and approaches, but because people have come to expect evaluations to look and feel a certain way: a clearly defined intervention, quantifiable impact metrics that are observable in the short-term, and experimental methods that control variables to isolate causal pathways. These are hard to get in system change work. We reflect here on five reasons why it is hard to evaluate a system change effort in these more familiar ways.

Systems don't change fast

Systems do not change on the timescale of a two-year funding round, or term in office. In fact, it is more likely to be decades[1] before you see lasting shifts in the way the system operates if you're trying to affect large scale social change. The more ambitious and wide-reaching your goals are for changing a system, the less likely it is that you will be able to see the fruits of your labour in the short-term. Funders who are used to receiving evaluations that demonstrate impact within the timeframe of a funding round, or a term in office, might have to shift their expectations about how long it will take before system change efforts realise their goals, and recognise trade-offs between seeking nearer-term, visible impact, and pursing deeper, longer-term change.

For example, anti-smoking campaigns that started up in the 1950s in the United States eventually reduced smoking rates by more than 60% among U.S. teens and adults, but it took fifty years to achieve this, thanks to decades of research funding, painstaking consensus building and widespread mindset and behavioural shifts on the part of the general population.[2]

In our experience of working to change systems, evaluating such efforts can be challenging, not necessarily because we lack the right methods and approaches, but because people have come to expect evaluations to look and feel a certain way.

System outcomes are often emergent

Emergence describes when changes happen at a collective level through the interaction of multiple elements in a way that is not predictable from examining the elements individually – making it hard to predict when and how changes will come about. For example, journalists and commentators closely tracking the gradual liberalisation of the USSR in the late 1980s did not predict that individual acts, behaviors and reforms would add up to the sudden collapse of the Berlin Wall on 9 November 1989, even if they might have had a sense that things were shifting. Instead this outcome seems to have been a result of an interaction between many different factors in the system that pushed it to a sudden tipping point.

In practice: We have seen how emergence can play out in our work on Drive – a national partnership addressing gaps in the system around high-harm perpetrators of domestic abuse. We worked with our partners SafeLives and Respect to develop a new service model, and advocate for supportive legislative change for five years, before a tipping point in 2019 saw our collective efforts come together in a wave of supporting legislation, funding and attention on working on perpetrators from central government. Evaluators would likely have had a hard time predicting how close work was to achieving these significant milestones had they been evaluating our efforts even 12 months before.

"Due to the sheer volume of interactions, factors, actors and entities… it can be difficult – even impossible – to identify clear and stable 'causal pathways'

Causal relationships are 'foggy' in complex systems

It is hard to track the links between specific inputs or actions and specific outcomes when doing system change work. Due to the sheer volume of interactions, factors, actors and entities in a given social system, and because of emergence, it can be difficult – even impossible – to identify clear and stable 'causal pathways' between the actions and inputs of system change effort and the outcomes it is seeking to achieve. It would be hard, for example, to say that the lobbying actions of a specific LGBTQ organisation solely **caused** marriage equality in the UK, even if they were instrumental. Instead, it was likely a combination of efforts by advocacy organisations and broader societal and cultural shifts that caused the change. Although we often expect evaluations to be definitive about what precisely an intervention can claim credit for, or what specifically caused a given outcome, in system change work, we need to accept more uncertainty, settle for plausibility and expect explanations that involve combinations of contributing causal factors.

The nature of the intervention is dynamic

A good system change effort will likely adapt, change shape, and re-route many times as it learns more about the system it is intervening in, and as the system itself changes. For example, an effort to radically reduce consumption of animal products may begin as an advocacy campaign, seeking to change people's attitudes toward animal cruelty, but turn into a food technology accelerator after identifying that developing cheap, tasty and convenient vegan alternatives is a more effective route to change than shifting consumer values. Alternatively, you might begin as a single organisation, but over time see the 'effort' as made of many partner organisations with complementary goals. For example, the Canadian Partnership for Childrens' Health and the Environment (CPCHE) began as individual community health groups, women's clinics, and childcare agencies, before coming together as a system change partnership with a common goal and overarching strategy to effect change.

These examples pose a challenge for evaluation both because they make it harder to know what exactly you're evaluating – what combination of initiatives, stakeholders, projects, people etc. constitute the 'effort' – and also because they make it harder to 'hold it constant' over the lifetime of the evaluation, meaning its scope might need to change half-way through.

There are multiple views on what constitutes both the system and success

Different actors in a system may articulate success very differently, making it hard to know whether they are aiming for the same thing – and hence to be clear when you have achieved it.

In practice: In our work on the Changing Futures programme – a £64 million initiative designed to change the way local services support people experiencing crisis – we found that the definition of success varied by region, sector and perspective on the system. For example, for some statutory partners, 'success' was the implementation of jointly commissioned support services that provided better coordination across hospital, mental health, treatment, police and housing services. For some community groups, 'success' was a reduction in the number of people in a given region falling into preventable crisis. And for others again, 'success' was a world where people who felt let down by services and society felt seen and heard.

Where we might be used to seeing success defined in evaluations in a relatively straightforward way, which signals an objective 'tools down' milestone, system change efforts usually demand a richer, potentially more complex definition of success, which may include some tension and contradiction.

A good system change effort will likely adapt, change shape, and re-route many times as it learns more about the system it is intervening in.

Evaluation is dead! Long live evaluation

Four ways evaluation needs to look and feel different in system change contexts

Despite its challenges, evaluation remains a powerful engine for driving forward system change work. The iterative nature of the work, its long time horizon, and the fundamental uncertainty of intervening in complex systems make it especially important to embed structures and mechanisms for deeply and continuously understanding the value of what you are doing, and for learning how to do it better. However, commissioners and funders need to shift their expectations about what the scope, timeframe and outputs of an evaluation will be, and system change practitioners need to think creatively about how to embed the learning and insights from evaluation activities into their strategic planning and their vision for impact.

A lot of thinking has already been done on what funders, practitioners and evaluators might need to do differently in the system change context, and in our work in recent years, we have seen a growing and thriving community of practitioners grappling with the challenges together. From Ashoka's focus on the iterative and interactive nature of system change efforts,[3] to Mark Cabaj's work on how to integrate system thinking and system-level results into your evaluation framework,[4] evaluators and 'doers' alike are sharing and testing approaches to evaluation that suit system change work. Among these, there are four ways that look and feel different in the systems context that particularly resonated in the context of the work we do.

1) Respect the 'black box' between inputs and outcomes.

Many seem to be searching for a 'holy grail' methodology that could, for example, replace randomised control trials as a widely agreed gold standard for conclusively demonstrating impact to funders and commissioners. It is tempting to want new methodologies that help us pinpoint precise, quantifiable results about the precise impact you get for your funding – say, an 0.025% decrease in prison entries for every £1k you spend on a crime prevention programme.

However, given the foggy causal relationships that exist in system change work, it is important for those commissioning system change evaluations to accommodate uncertainty and accept the limits on proving or comprehensively 'uncovering' causation. This might look like tolerating lower levels of certainty about what 'really' caused an outcome, or being willing to accept multiple possible explanations for why something happened the way it did. Instead of expecting to map complete, precise 'causal paths' between an intervention and given outcomes, system change evaluation should seek to sharpen thinking at the margins of the unknown, observing patterns and clusters of inputs that together seem to cause an output, even when the exact relationship between the variables isn't clear and likely involves a great deal of variation.[5] But this is not to say that evaluators should abandon causal **thinking** – trying to understand what happened **and why** is central to what makes evaluative work valuable and

actionable – just that we need to broaden our toolkit of approaches, accept more uncertainty and work with wider, looser and less complete hypotheses.

2) Embrace embedded evaluation.

We are used to the idea of independent evaluators, who stay at arm's length from the work, and clearly distinguish themselves from the 'doers' – only engaging at pre-determined points for fieldwork, and seeking to observe 'at a remove', rather than establishing an ongoing dialogue with practitioners. Maintaining an evaluative perspective is important; as HMT's Magenta Book says, to be credible, an evaluation should aim to 'remain independent and unbiased'.[6] But as we have discussed above, in system change contexts, the ever-changing and dynamic nature of both the system and the work itself as it navigates and adapts makes it less possible, as well as less useful, for evaluators to stay entirely detached.

Instead, to make meaningful and informed judgements about the value and impact of the work, and to keep up with (and contribute to) new learning and course-corrections, they need to be embedded more in the work. We see increasing interest among those working at system change in evaluators getting their hands dirty as 'part of the team', sharing real-time feedback and being involved in the development of strategy,[7] or growing demand for 'learning partners', where the emphasis is placed on unpacking and exploring impact in context and applying it in actionable ways with stakeholders. Embeddedness might also come in the form of an expectation that practitioner teams can demonstrate their capacity to hold 'evaluative thinking' in sufficient tension with their strategic vision for the system – i.e. that they retain a willingness to reach the 'unwanted conclusion' that perhaps what they are doing is not having the impact they want it to.

3) Orient the work in the wider movement of change.

The elements of a given system change effort should not be considered in isolation from their wider context. For example, Changing Futures – a large scale government-funded programme to facilitate local innovation in delivery of coordinated services for those experiencing multiple disadvantage – is aiming to shift accountability in the statutory sector for keeping people from falling through the cracks between services, catalyse funding reforms, and spread adoption of new frontline case work models. But this is not the only big intervention aiming to transform outcomes for this cohort. Arguably, coalitions like MEAM and Fulfilling Lives[8] have been laying the groundwork for other 'building blocks' – new public conversations and new ways of working. System change evaluation – and those who commission it – should be concerned with how these efforts come together to shift the system, not just how they work in isolation.

There are also other events in the wider system that are not deliberate contributors to a given system change goal, but that are advancing or hindering it in meaningful ways, which system evaluation should seek to understand – from rising waves of new public interest in response to geopolitical or cultural events, to the rapid uptake of new technologies. Our Routes to Scale framework offers some prompts for thinking about changes that might be taking places across the system that you could take into consideration. There are many others too – including FSG's water of system change,[9] Donella Meadow's leverage points,[10] and the System Innovation Initiative's Four Keys.[11]

Instead of expecting to map complete, precise 'causal paths' between an intervention and given outcomes, system change evaluation should seek to sharpen thinking at the margins of the unknown.

4) Consider the quality of 'inputs' into the work, too.

For many good reasons, over the past few decades there has been a shift away from input-based performance monitoring toward outcomes. This is in recognition of the tendency of an inputs focus to stifle innovation and distort programmes in favour of process fidelity at the expense of achieving the desired impact, typically holding the 'doer' to a set of pre-determined activities rather than unleashing them to achieve outcomes by whatever means are most effective.[12] However given the long timeframe over which visible results might be attained in system change work, and the difficulty of properly observing and understanding them when they do, paying attention to how well set-up a system change effort is for its task is arguably a valuable subject of evaluation too.

A good way of thinking about this is the analogy of a cancer lab that has been working on a cure for cancer for decades, without success, and little to show that they are close. Despite the lack of the desired outcome, it's not useful to think of the lab as having failed; instead to help evaluate the lab's work, you could look to the quality of the 'inputs' into its setup. These might include: its adherence to the scientific method and medical research protocols, the calibre of the researchers, its ability to learn and adapt and the strength of its governance.[13]

This isn't to abandon outcomes – 'failed' research about what **doesn't** cure cancer is an outcome, too, and the lab should be narrowing the field of possible solutions as it learns, chipping away at the 'black box' – but when achievement of end goals are distant, 'inputs' are a helpful subject of evaluation too. To put this in the system change context, some of the inputs that are likely most important for ensuring a system change effort is adequately equipped to achieve its goals include:

- capacity to inquire, learn, reflect about the wider system and nature of impact
- its willingness to acknowledge failure or 'dead ends' and re-orient appropriately
- its embeddedness in the wider system and legitimacy with key stakeholders
- quality of its governance and leadership.

Not throwing the baby out with the bathwater – three ways evaluation will stay the same

Although funders and commissioners may have to accept evaluations that look and feel different in the world of system change work, at its core, evaluation still serves the same function: making a judgement about the merit, worth or significance of a piece of work by combining evidence and values, as a result of systemic process.[14] This task is no doubt much harder in system work, but for that reason, some foundations of evaluation must remain (and even be fortified) as we venture into the system change context.

1) System evaluation should remain directional – but strike the right balance between a focus on learning and one on results

Traditionally, funders have expected evaluators to focus on understanding and evidencing the impact of programmes based on a set of outcomes and indicators that reflect success, agreed in advance with stakeholders.[15] Impact evaluation in particular has focused on understanding whether a given intervention had a positive impact, demonstrating that intended results follow from programme results whether directly or indirectly.[16] The evidence-based policy movement (EBPM), from which this grew, had its roots in medicine and its tradition of using experimental pharmaceutical trials (usually Randomised Control Trials) to show the effect of specific treatments. It emphasises that policy should be informed and able to demonstrate and where possible measure 'results', 'value for money' and 'effectiveness' – and impact evaluation has traditionally been the tool used to demonstrate that a given policy 'worked'.[17]

There are clear risks in applying this approach uncritically to system change efforts. System work is defined by unstable and dynamic contexts, complexity and interacting components, a high degree of uncertainty about what the right course of action is, and, sometimes, a changing view of success. An evaluative approach that only admits the precise empiricism of 'classic' evidence-based policy methods risks supporting only high-fidelity and (arguably) controlled interventions: discouraging work on broader, system-level approaches that may be more 'emergent', a clear case of the tail wagging the dog.

At its worst, this approach can risk a myopic focus on those outcomes which can be achieved and observed in the short-term, neglecting the more subtle shifts that might be creating the conditions for transformative change down the line.

Is there any role for classic experimental approaches, like Randomised Control Trials (RCTs), in system change work? We have outlined the ways in which experimental methods – methods that try to isolate the approach under test from other variables, require clear timelines and an intervention that is put in place according to the test protocol, and seek a high standard of certainty to identify causal factors by comparison to a counterfactual – conflict with the scope and complexity of system change work. However, while there is little scope to apply such approaches to the broad, dynamic shape of 'whole system' work, this is certainly not to say that such methods cannot be useful in an effort to change a system. For example, one might identify that an important part of an overall system effort to ensure every child feels safe, empowered and supported online is to identify school-level interventions that are effective in reducing social media bullying. Finding interventions that are rigorously tested and can be shown to be effective might sit alongside the need to strengthen regulation of social media companies, build parental awareness and shift educationalists' mindsets. An RCT or quasi-experimental method could be the most appropriate and effective way to identify these interventions and build a solid evidence case, feeding into and sitting alongside other parts of the work. Evaluators have also looked for ways to apply RCTs outside of specific interventions – for example, in designing them around the **principles** to be enacted through an intervention, rather than a strict protocol of pre-determined activities.[18]

Recognising this, one set of approaches to evaluation in system contexts emphasises that the emergent quality of outcomes in system change work, and the inherent 'unknowability' of systems, make tightly articulated hypotheses for how change will come about meaningless. Instead of strategising for change, proponents of this position argue that the focus should be on learning how the current system behaves, on self-reflection, mindset and culture shifts, and on nurturing intangible but important qualities such as empathy and trust. Progress, on this account, is made by a system becoming better at learning: this is the result to look out for and to build accountability around.

These approaches acknowledge the crucial need for evaluative work in system change contexts to remain flexible and open to shifts in both the end goals – which may change or become clearer as you progress – and the approaches for getting there. They rightly emphasise that an experimental approach that relies on defining fixed end outcomes (allowing no possibility of change or update) and controlling implementation to stick to the theory of change under test cannot work when dealing with the scope and complexity of systems. But there is also a risk here. Learning, too, needs direction. The cancer lab that has great learning practices about cell mutation, but does not link these to an ultimate goal of moving closer to finding a cure by narrowing the field of fruitful research routes may be learning – but it may not be progressing towards its end goal. Similarly, a system change effort focused on early childhood education may generate a considerable amount of interesting insight about how to improve the relationship between schools and social workers, but unless it keeps sight of the hypothesized connection between this and child outcomes, and is set up to change direction if none appears to be found, it risks focusing learning on an interesting but ultimately marginal topic. If we are reticent to name – even loosely – a destination, how can we know if we are heading in the right direction?

The instinct to rigorously check for intended impact is well-placed, as good intentions do not always lead to good outcomes. A focus on results helps funders and practitioners focus on impact rather than becoming rigid about the processes and activities used to achieve them, and also helps distinguish between work that 'made the difference' and work that might have been good or useful in other ways, but did not in the end matter for achieving the desired outcomes. This prevents would-be changemakers from 'spinning the wheels' around outcomes that would have been achieved anyway or sinking resources into approaches that are going nowhere.[19]

In the systems context, staying focused on results is harder, not least because the 'end' goal results are likely much further off, but perhaps also because they are more loosely defined, or more multi-faceted. But arguably it is even more important to stay focused on what change you want to bring about when the change is harder to realise. Although new learnings, unexpected consequences, and exogenous shocks might change how you think about your 'end goal' results, or which early indicators of them to focus on, clearly articulating a destination, and checking progress against it remains the most reliable way to get there, and the best way to prevent waste or unintended harm. Further, given the need to iterate and learn over time, focusing on results is a vital tool to help prioritise what you need to learn **about**. While there may be some points in a system change journey that demand complete openness to new information (e.g. right at the beginning, before you know anything about the system, or after a significant exogenous change), learning without a clear goal is at best a waste of scarce resources, and at worst, can lead you off course by encouraging learning about what is interesting or salient, rather than what is most instrumental in achieving real impact.

Although new learnings, unexpected consequences, and exogenous shocks might change how you think about your 'end goal' results … clearly articulating a destination, and checking progress against it remains the most reliable way to get there.

Why is it so important to check what our impact actually is? Our best guess for what works, or instincts about what 'must surely' be the right approach is often incorrect. A good example of this is so-called 'scared straight' programmes. These programmes were a popular strategy to reduce the chances that children at risk of involvement in criminal activity become involved in the criminal justice system. The programme involved organised visits to adult prison facilities by incarcerated young people or those at risk of incarceration, where adult inmates shared graphic stories about life in prison, to dissuade young people from criminal activity. Multiple evaluations found that this programme likely had the opposite effect to that intended, with one report estimating that reoffending was 68% higher among young people who participated in the programme compared to those who did not. It always pays to check our assumptions.[20]

2) System evaluation should retain a set of standards for rigorous inquiry

Alongside directionality, we should also be reticent to depart from the idea of rigorous standards of knowledge in the context of system evaluation, even if we must make concessions to complexity and uncertainty. Zenda Ofir argues that evaluation is both art and science; a discipline which demands intuition, empathy, and reflection on lived experience, but also critical thinking, technical acumen and reasoned judgement.[21] In standard impact evaluation approaches (examined above), the 'science' of evaluation is arguably privileged above the 'art', and is safeguarded by a set of strict standards of independence, methodological transparency, and hierarchies of evidence. This has often kept evaluators at arm's length from the work itself, and able to clearly differentiate their role and perspective from that of the 'doers' – those responsible for making strategic decisions, and setting a course of travel.

These ways of distinguishing evaluators' role tend to blur in system work; where evaluators need to be more embedded in order to keep up with and contribute to new learning, and truly understand the different dimensions of impact. As Julia Coffman outlines,[22] the role of the evaluator has rightly expanded from its roots as 'applied social science researchers' to encompass work that we might see as belonging to 'theorists', 'strategists', 'strategic communicators', 'systems thinkers' and 'facilitators, coaches and trainers'. However, to hold onto the role as 'critical friend', able to test and challenge implicit hypotheses, surface values and assumptions, and reach uncomfortable conclusions ('perhaps this isn't working?'), system evaluation needs to retain at its heart, even as its scope expands and 'loosens', a commitment to a certain kind of (scientific) rigour: to critical thinking, reasoned judgement and technical acumen (Ofir[21]). Drawing on and informed by the many different statements of such principles,[23] we have summarised these as: transparent reasoning, rigorous methods, reflexivity, commitment to causal reasoning, and fallibility (see Table 1 below).

Undoubtedly, these principles can be (and have often been) applied to system change work in a way that is ill-fitted to the task of understanding the ways a system is shifting, and the role of specific effort within that. This is particularly the case for the use of experimental methods. However, we think it is the expectations and practice, rather than the **principles**, that have caused issues in system change work. The principles may look and feel different in the context of system change work – but their core goals remain unchanged.

Evaluation is both art and science; a discipline which demands intuition, empathy, and reflection on lived experience, but also critical thinking, technical acumen and reasoned judgement.

Table 1 – Principles of evaluation

Principle	How it is conventionally expected to look	How it may look in system change work
Transparency of reasoning Allowing others to test, engage with, replicate and/or confirm embedded hypotheses	Fully articulated set of hypotheses producing clear & complete evaluation scope that is held constant over time (even if the needs and nature of the work have changed)	A clear 'system view', system boundaries are articulated, and hypotheses of change within the system (which may be partial) set out and updated over time based on learning; iterative evaluation scope that evolves and is updated as appropriate
Rigorous methods Methods that are appropriate for the given questions and can be clearly communicated and (in theory) replicated	Experimental methods preferred and highest levels of certainty are sought; RCTs seen as gold standard, other methods sometimes distrusted	Wide range of appropriate methods are deployed for the questions being asked; experimental methods used in targeted ways where warranted
Reflexivity and limiting bias Self-awareness about biases and role of evaluator in shaping assumptions and findings	Independent evaluators working separately from the project team; demand on, and confidence in ability of, evaluators to generate truly objective findings	An embedded evaluation partner 'checks and challenges' from 'within' the project work and in real-time; awareness of bias and surfacing of different perspectives, but still shaping a shared interpretive space
Causal reasoning Clearly articulated hypothesised relationship between intended actions and their effects	Causal paths can be comprehensively mapped and experimental methods or those focused on attributing impacts to specific actions are elevated above other (e.g. non-experimental or qualitative) methods	Ambition to sharpen and refine hypotheses as you learn in order to inform next steps, while respecting the 'black box'; thinking in terms of contribution rather than attribution
Fallibility Willingness to be wrong about the value of a given piece of work	Establishing a clear, compelling and testable counterfactual considered the only acceptable way to test hypotheses	Openness to being (and willingness to be) wrong, seeking out the 'undesired truth' (i.e. that the approach is not having the desired effect) about the work, and maintaining integrity of voice

Isn't causality just too hard to get at in complex systems? It is tempting to argue that, rather than asking funders and commissioners to relax certain expectations, we should push them to abandon them altogether. If we can't conclusively demonstrate impact and attribution in system change work, maybe we just need to let go of our desire to understand causation in these complex contexts?

But that would be a mistake – and one that some are already making. Julia Coffman of the Center for Evaluation Innovation, warns that 'evaluators of complex philanthropic strategies often do little to unpack assumptions about what happens in the black box of change once strategies are unleashed', focusing instead on 'describing observed changes without investigating how or why they occurred' or – even more problematically – 'simply assum[ing], without investigation, that relationships exist between implemented strategies and observed outcomes.'[24] This risks that we continue allocating scarce resources to approaches that are not bringing about our end goal impact, or that we let the loudest voice, rather than the best strategy, dictate our direction.

Rather than abandon causal analysis along with the narrow focus on experimental and quasi-experimental methods, system change evaluation should maintain a focus on causal thinking, even if the methods for exploring it and the levels of certainty it offers are different. System change evaluation has a vital role in making the most impactful pathway to change brighter (shining a torch into the 'black box'), even if it can never reveal and prove it beyond a doubt. Rigorous standards of inquiry and commitment to learning more about causal pathways remain our best approach to sharpening our system change strategies at their margins, testing hypotheses and generating new insight about the best strategy.

3) System evaluation should not give up on the goal of mediating between different views to reach a conclusion about the value of a piece of work

As we have set out, practitioners and funders must grapple with the challenges of how to approach methodology and evidence in the context of system change evaluation, where common 'go-to' tools and familiar standards may not be available. But there is a third core element to evaluative work which also calls for consideration: the role of values. Evaluation is 'any systematic process to judge merit, worth or significance by combining evidence and values'.[25] We have considered how 'systematic process' might look different in a system change context – and how thinking about evidence can differ when it is far-off 'system outcomes' that you ultimately seek to measure your success against: but how do evaluators deal with the question of which value(s) to combine with evidence to come to conclusions?

The scope and complexity of system change work necessarily means that a wide range of actors are involved. They may bring differing value sets from which to judge the work's success, including differing perspectives of the structure and boundaries of the system, different views on what goals are valuable, on the value of evidence gathered and on its implications for future decision-making. Faced with such pluralism, it can be tempting to simply note the differences and give up on the task of synthesising or reconciling different value sets. The risk of this, however, is either stasis – it is hard to move evaluative work forward without agreement on the values against which conclusions are to be drawn – or a situation in which the loudest voice wins.

Instead, those leading evaluative work have a role in navigating these differences by establishing a shared deliberative space with agreed 'rules of engagement', in which different perspectives can be exchanged and debated based on agreed, shared parameters. Evaluative practice should work at surfacing the values and agendas held by different stakeholders, drawing out underlying assumptions, making these clear for all,

articulating any conflicts they observe and encouraging stakeholders to work together to come to conclusions. For example, work focused on reducing vandalism may include groups pushing for more 'restorative justice' approaches, and others arguing for clearer and swifter sanctions. Part of the evaluator's role would be to understand whether both groups, despite appearing to lean in different directions, in fact agree on an overall goal of reduced vandalism, and are prepared to follow whichever route the evidence favours, or whether the different approaches are driven by more fundamental values – seen as 'right' regardless of their relative impact on vandalism rates– in which case, there may be a need to recognise two independent and different system change efforts with different aims. As Emily Gates puts it, the idea is not to provide the answer, by selecting the 'best option', or by ascertaining the set of values with the most support, but to support the process by seeking to understand the consequences of differing value interpretations, setting these out, and encouraging generative deliberation.[26] This is a greater ask on evaluators than they may be accustomed to, moving beyond assessment of evidence into engaging with the values that shape the motivation(s) for system change work and the theories about what can bring it about.

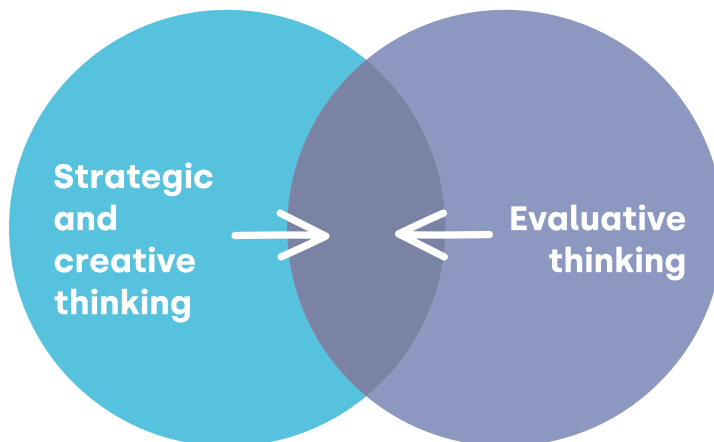
In practice, this might require from funders, commissioners and decision makers an ability to clearly articulate values-based decisions and their rationale, including distinguishing between instrumental beliefs ('we're funding/pursuing this because we think it will achieve X, which we value – but we will change strategy if it doesn't') and intrinsic ones ('we're doing this because we value it in itself, regardless of what other impact it has and where it leads').

Evaluative practice should work at surfacing the values and agendas held by different stakeholders, drawing out underlying assumptions, making these clear for all, articulating any conflicts they observe and encouraging stakeholders to work together to come to conclusions.

System change evaluation in practice

If you are setting out on a system change journey – or funding one – how should you think about your impact and whether you are making progress? What does it look like in practice to balance some of the newer approaches – respecting the 'black box', embedding evaluation, orientation toward a wider system change movement – with some of the 'core' roles of evaluation in checking for results and rigorous causal thinking? We suggest keeping in mind the following five things when planning your learning and evaluation approach:

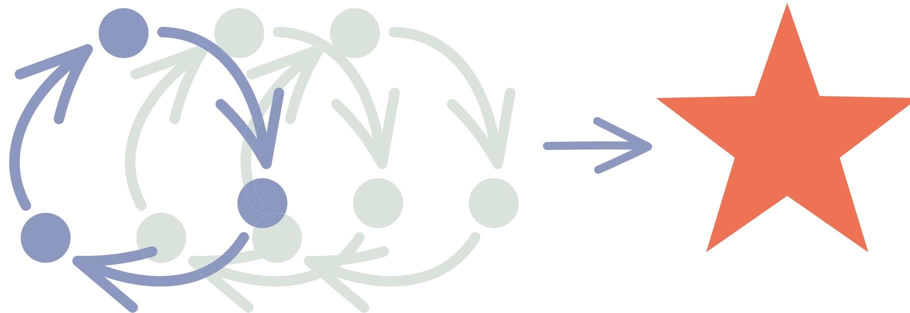
Embed evaluative thinking into your strategic planning



In system change contexts, it makes less sense to separate evaluative thinking from strategic planning. Expect your strategy to need regular iteration as you learn more about the system and your impact in it, and plan to hold evaluative thinking – surfacing and testing your assumptions, understanding your biases, considering your wider system impact, dispassionate appraisal – in productive tension with your strategic and creative vision.

Independent assessment will always be highly valuable. But you can achieve it in a few different ways. Embedded-but-independent evaluative thinking could mean you create an in-house 'evaluation and learning' function with responsibility for checking and challenging your strategy and vision, and integrating learning as you go, or that you build in regular time with strategic leaders and decision-makers to wear an 'evaluative thinking hat' rather than a strategic one. You could also bring in a learning and evaluation partner to offer a regular 'critical friend' voice to your strategic planning, and establish learning feedback loops to ensure you are adapting and course correcting as needed.

Plan for iteration, but don't lose your 'North Star'



While it is important to expect adaptation and iteration as you learn, it is equally important to have a 'North Star' to act as a compass for the work. Articulating what you are trying to achieve and how you think you might achieve it, even loosely and with known (and unknown) knowledge gaps is an essential step for ordering the information you will receive about the system, and for learning in a way that helps you refine and strengthen your strategy. As new information, an ever-changing system landscape, and unexpected outcomes (and very likely, failures) shape and update your thinking, you'll need to regularly go back to the drawing board, and potentially re-route entirely. But having a clear goal (and hypothesis for how to bring it about) will ensure that this is a process of refining and sharpening your thinking about how to reach your goal, rather than constantly starting from scratch. Our Routes to Scale framework offers some prompts for thinking about what shape your goals for system change could take and what your first hypotheses for bringing them about might be.

In the example of an effort with a 'North Star' of reducing animal cruelty caused by factory farming, iteration with direction might look like:

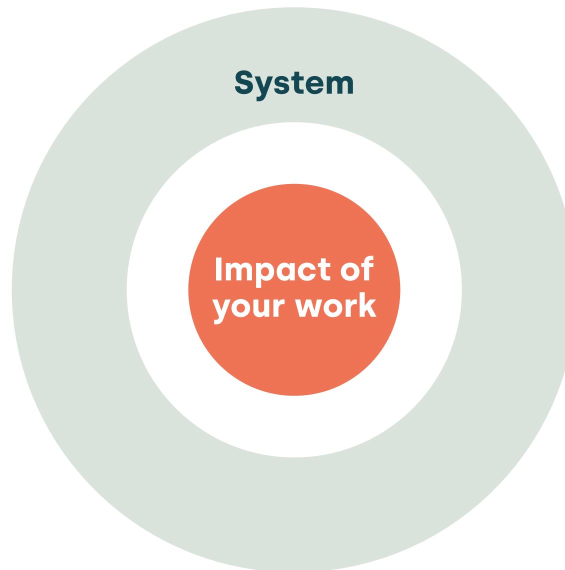
(1) Articulating a goal to reduce consumption of animal products, and implementing a strategy to change attitudes toward animal cruelty by making animal suffering more salient.

(2) Learning through experimentation over time that a much faster route to change is to stop appealing to people's values (hard and expensive to shift!), and instead to invest in accelerating the production of cheap, tasty and convenient lab grown animal products.

(3) Re-stating your goal to be about reducing dependence on factory farming (rather than reducing consumption of animal products), and adapting your strategy based on what you've learnt.

In this example, the 'North Star' has remained a compass for the work, even as the strategy to get there and the way you articulate your goals have changed over time.

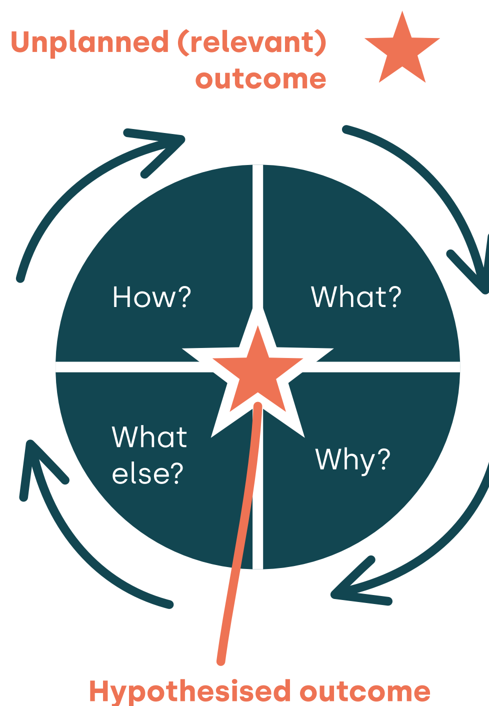
Check for impact regularly, but put it in (system) context



Demonstrating impact remains important in system work. However, to avoid a narrow focus on intended results, it is important to be sensitive to what is happening in the wider system because of, despite, or adjacent to your effort – both good and bad. We advocate for rigorous reflection about what the results mean, what assumptions they are dependent on, and about the interaction between your specific effort and the wider system: did your effort interact with the wider system in an unexpected way? And what else happened in the wider system that you might not have accounted for?

For example, as part of a system change effort to widen access to meaningful, lifelong employment, perhaps you have prioritised developing and then scaling up job coaching services in boroughs across London for those leaving prison. As a result, you have got thousands of people leaving prison into lasting employment they would not have otherwise been able to attain. However, a scan of the wider system reveals that rather than expanding access, this programme has actually diverted funding from employment services for other groups at-risk of long-term employment. To achieve your underlying system change goal, you consider ways to adapt your programme to widen referral pathways beyond prison services, and expand the scope of the service in a more cost effective way.

Retain rigorous, causal thinking but be sensitive to emergent outcomes



As explored above, part of making a judgement about the value of a given piece is establishing not just what happened, but why it happened, and what (combination of things) might have caused it. This is as important as ever in system work. However, given what we have discussed about emergence – that outcomes may emerge in ways you could not predict or account for in advance – causal thinking in system change work should be tempered with a sensitivity to emergent outcomes, and make space for things to occur in ways that a specific effort could not have solely 'caused' (even if it contributed). In practice this might mean that funders expect evaluation to unearth a story that blends 'things we can reliably show we caused', with 'reflections on ways the system is changing or outcomes are emerging' alongside our work.

Homelessness charities in the UK have been working tirelessly for decades to end homelessness through a mix of advocacy, awareness raising, campaigning, and direct financial support. However, in early 2020, at the outset of the pandemic, extraordinary action was taken across the country to move everyone at risk of street homelessness into safe accommodation; rough sleeping in England fell by 37% between autumn 2019 and autumn 2020. No one working on the effort to end homelessness could have predicted COVID or its impacts on homelessness; but this exogenous shock shifted the system into a new equilibrium almost overnight. Although homelessness has subsequently risen, the pandemic shifted expectations about what is possible.[27]

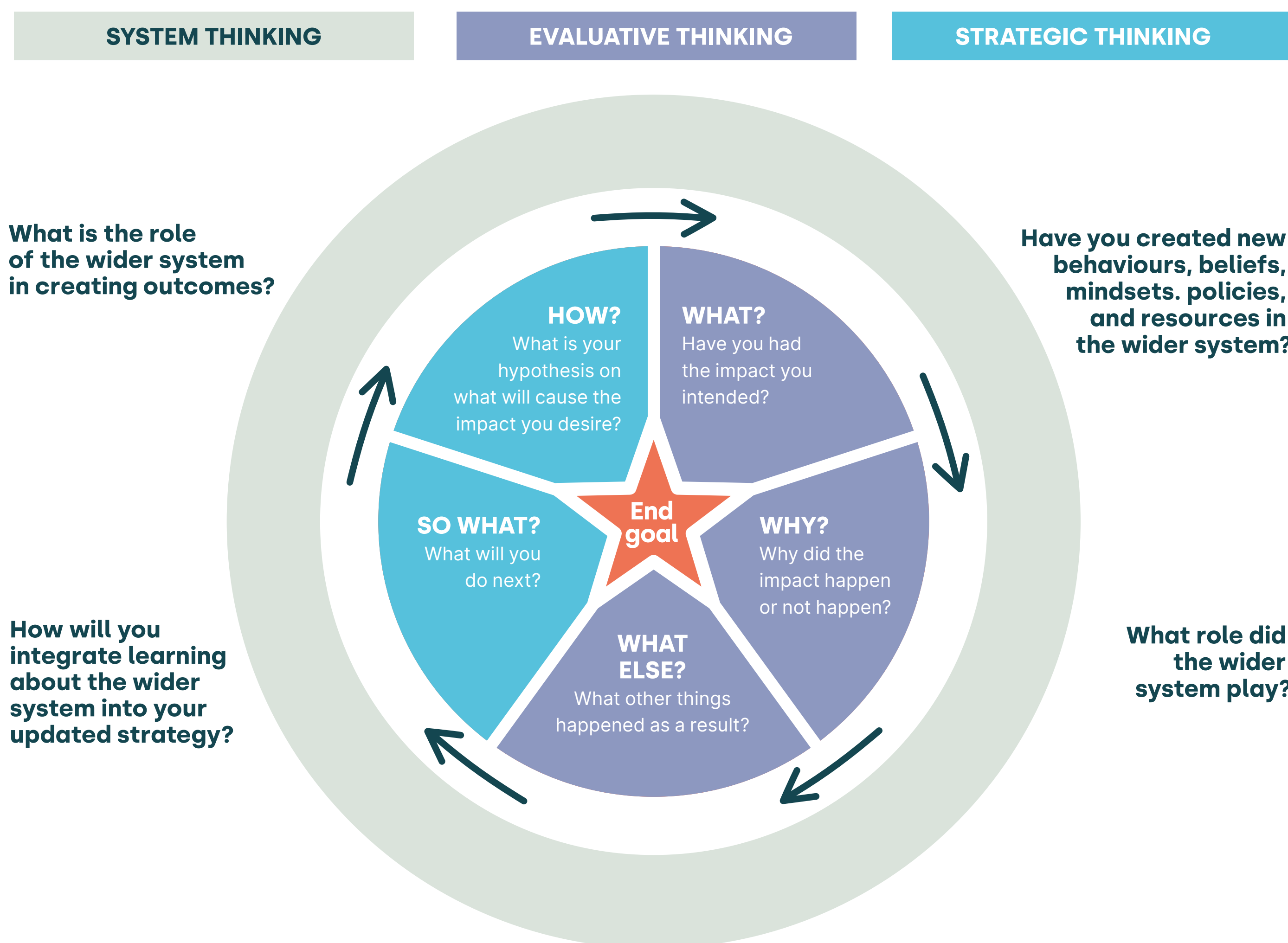
Treat your strategy as a hypothesis and learn with direction

Henry Mintzberg says that you don't plan a strategy, you learn it. This is especially true of system change work. Although it is tempting to think that good planning and analysis will set up you up with the 'right' strategy, and allow you to transition from 'planning' to 'doing', thinking evaluatively about system change efforts demands learning and iteration as the status quo.

Learning in a way that helps you refine your strategy is hard to do well; it requires building appetite for challenge, inquiry and change among stakeholders, constant questioning of assumptions and willingness to be wrong. But without providing the structures, culture and mechanisms for constant learning, system change evaluation is unlikely to help drive the effort forward.

Cycle of learning questions

We have tried to bring these ideas together in our **cycle of learning questions.** Starting with defining – however loosely – your end goal, the cycle works through a set of questions that encourage you to set your starting hypotheses, take a wide lens to search evidence of impact in the context of the broader system, explore potential causal links between changes and your work, and feed this back into your strategic planning:



How to know you're making progress? An inquiry framework for system change

We are interested in how evaluative thinking in a systems context can make space for learning and emergence alongside strategic direction and the importance of holding ourselves accountable for achieving impact. A challenge in system change work is that it is hard to articulate the changes you want to make; it's tempting to either reach for very high-level statements (think 'a healthier education system') or zoom in on narrow or specific changes.

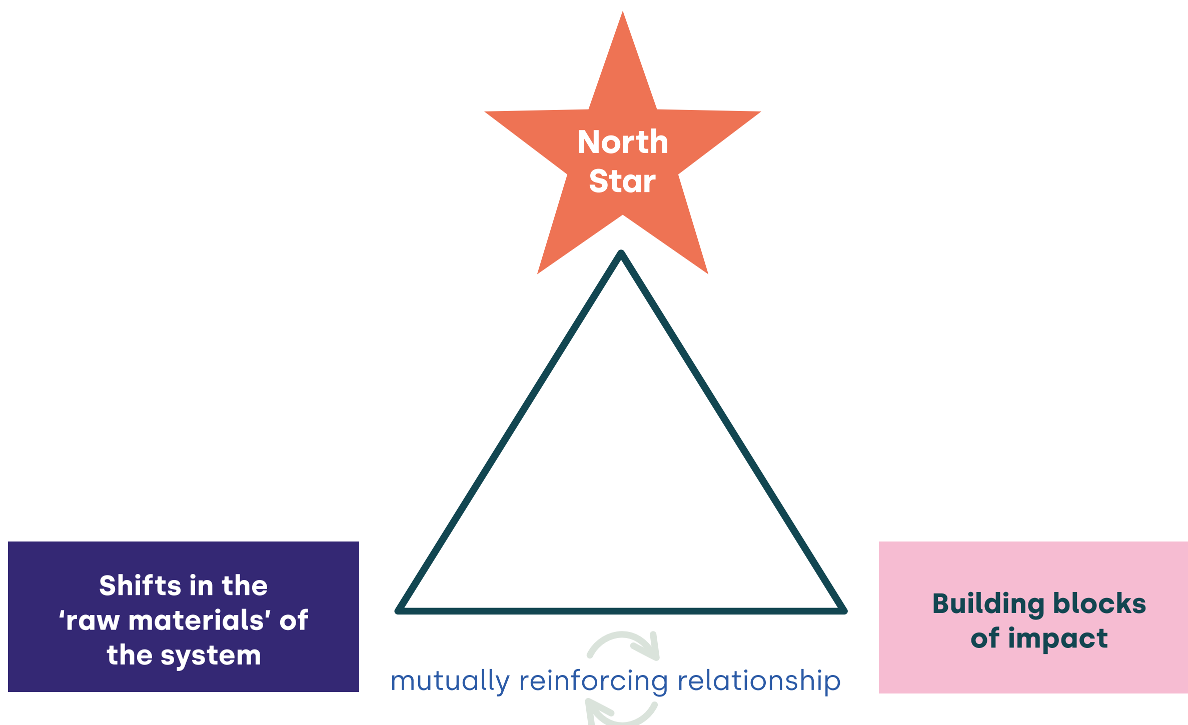
We think there is a case for thinking more about what can bridge the gap between this high-level appreciation that the system must change to generate new outcomes, and the individual 'theory of system change' particular to each effort. Similarly, during our engagement with practitioners, evaluators and funders of system change work, we heard a clear call for evaluative approaches that help pick up the signs or signals that the system is shifting, and models for thinking about whether these early indicators will lead to deeper, more sustained changes over time. We see an opportunity to build an evaluative framework that links:

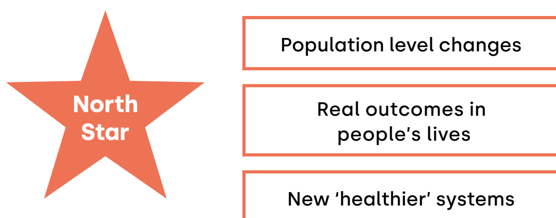
- what it takes to create the 'conditions for emergence' in the short- and medium- term (before longer-term transformation might be visible or possible)
- a clear sense of direction and progress towards manifest changes in the way the system is configured, and
- tangible outcomes that show how work is making lives better

There is no stand-in for the hard, ongoing work of thoroughly understanding each system we seek to change, but based on more than a decade's experience of working with partners to change systems to achieve better outcomes for people, we think there are core things all impact hypotheses should consider. Our **triangle of system evaluation** can be thought of as a set of guardrails for evaluative thinking, a structure to order thoughts, and a way of articulating what sort of changes you might need to look for to know whether your system change work is making progress.

A challenge in system change work is that it is hard to articulate the changes you want to make; it's tempting to either reach for very high-level statements (think 'a healthier education system') or zoom in on narrow or specific changes.

The triangle of system change evaluation





The 'North Star' provides the direction towards which a given system change effort is oriented. It may be a crisp articulation (e.g. 'eliminating knife crime in Glasgow') or a broader sense of direction (e.g. 'a healthier Costa Rica'). Although it is a loosely held guide that may change over time as you learn, it ensures that shifts in the system materials and the establishment of new sustainable configurations are **directional** – that is, oriented towards and pursued in the service of a vision of what a new system or world would look like – the change you are trying to bring about. While there should always be hypotheses about the connection between the North Star and a shift in the system materials or a building block you are seeking to put in place, these may sometimes be high level and vague – and at other times may coalesce into a more precise causal relationship.

For example, you might start with a vision for 'a healthier Costa Rica.' As your familiarity with the health system in Costa Rica grows, you might re-cast your vision as being about 'a Costa Rica where health care priorities are determined by the needs of local communities, and health care is centred around strong relationships between health care workers and the communities they serve.'[28]

Shifts in the 'raw materials' of the system

- New values and beliefs about what's possible
- New culture and behaviours
- Power shifting to new groups
- New relationships between parts of the system

The second element of our triangle highlights the role of the 'raw materials' of the system in unlocking change. Drawing on the System Innovation Initiative's Four Keys, and FSG's Six Conditions of System Change,[29] we think of these 'raw materials' as the deep, embedded, or implicit patterns of the system. For example, mental models about what the system is 'for', widely used metaphors to describe the issue, where either formal or informal decision-making and agenda-setting power lies, or the quality and nature of relationships between key actors. These 'materials' – mental models, beliefs, culture, power, relationships – are often locked together in patterns that reinforce the way the current system operates. It is only once the existing arrangement of these materials is disrupted that new system configurations can emerge.[30] Shifts in the raw materials might be early indicators that you are making progress, or that the conditions for emergence of the more tangible changes you need are being established.

In our work as part of the Drive Partnership – a partnership between Respect, SafeLives and Social Finance to transform the national response to perpetrators of domestic abuse – it was our hypothesis that developing new ways of working with perpetrators to reduce domestic violence, and unlocking supportive national legislative reform would not be possible without first changing beliefs and mental models around domestic abuse from 'why doesn't she leave' to 'why doesn't he stop'.

Building blocks of impact

- Supportive policy, legislation, regulations
- New or changed funding
- Accountability or incentives shift
- New public conversation
- Sector adopts new approaches
- Widespread solution delivery

The third element is the building Blocks of impact, drawn from our 'Routes to Scale' framework. The building blocks can be thought of as the significant and tangible manifestations of change in a system that lay the foundation for lasting impact at scale. We think of these as significant milestones to track and measure progress toward system change. For most issues, multiple building blocks will need to be in place to create a solid platform for lasting widespread change.

The blocks are not elements of a single organisational strategy, but span the system, and can be put in place by many different actors who share a common vision over time – or emerge through unpredicted shifts in the wider context. Finally, the blocks are not in a fixed order, but are mutually supportive; making progress towards one can open windows of opportunity for change in others (e.g. a change in public conversation might have to precede legislative reform).

In the effort to improve the quality of housing for those experiencing the highest inequalities, Victorian-era philanthropists played an instrumental role in unlocking funding streams for housing societies by developing proof of concept initiatives and increasing the pressure to react to growing inequalities. This funding was the precursor to the UK's modern-day Housing Associations.[31]

This is not unidirectional: they are likely be in a constantly moving and iterative relationship – establishing a building block may open up the possibility of even deeper shifts in the system materials, and so on.

Together, the raw materials and building blocks provide guidance on what to look for to indicate that a system is changing in sustainable and lasting ways. Evidence that the existing pattern of the materials is being 'unlocked' and re-arranged will tell you that you are creating the conditions for change (e.g. new building blocks) to emerge, or that you are embedding new system patterns, and evidence of new building blocks provides a clear signal that the system is being re-arranged in tangible and sustainable ways.

The materials of the system and the building blocks of change should be thought of as mutually reinforcing. Shifts in the system materials can create the conditions for the building blocks to be achieved, by 'tilling the soil' of the system, allowing new configurations to emerge, and sustaining and reinforcing them once they have been established. In turn, the building blocks are the manifestations of the shifts in the raw materials, which codify and embed them into new, sustained configurations.

In 2013, the Australian parliament passed ground-breaking new legislation to shift funds for disability services from service providers to individuals living with disability, enabling them to spend on which ever service or product they think is most valuable to them in managing their disability (See NDIS). In a legal sense, this directly shifted decision-making power over care services to individuals living with disabilities. However, in practice the legal change has been meaningless without the disability sector adopting completely new ways of working, building new relationships between those living with disability and their network of service providers, and without a new mental model, supported by a broad new public conversation around what the purpose of the disability care system should be.

When brought together into a single framework for thinking about progress, the triangle provides some guidance for how to navigate the uncertainty of system change:

- **Maintain an orientation toward your 'North Star'** – that is, your vision for the end goal you want to achieve. This could be better outcomes for specific groups (better health outcomes for children experiencing disadvantage), population-level shifts (healthier children and young people), or less tangible impacts on the way the whole system operates (health system focused on prevention). It might be decades away from being realised, and the way you articulate it might change over time. But we think articulating it and holding onto it is core to achieving system change. Social Finance pioneered the Social Impact Bond, which in some ways is an archetypal approach to putting the 'outcome first'. We have refined and nuanced our thinking a lot over the years about how to define and measure outcomes since setting up the first social outcomes contracts, but have retained a belief in the power of articulating a clear goal as a rallying call for partners to mobilise around to achieve change, and an 'Occam's razor' that helps parse out what matters and what does not, to get things done.
- **A common language to describe what you're doing and why.** The triangle gives partners a common language for describing the different planks of their collective effort, and how they are going to achieve their goal. Using a common language helps partners to keep each other accountable for articulating the relationships between actions and the results they are achieving (or not achieving), and helps identify new or buried assumptions more efficiently.
- **A way to talk about short-term changes before you have anything tangible to show.** As Anna Birney notes, 'we could be doing something for 20 years and the system doesn't feel like it has shifted or tipped, but it doesn't mean that you're not creating system change, you're not creating conditions for emergence to happen'.[32] Our system evaluation triangle invites you to examine some of the early shifts in the system that could indicate new patterns are emerging, and help you to demonstrate how you might be creating the conditions for emergence, even if you can't show anything tangible yet.
- **A thought prompt for articulating your pathway to impact, and a set of questions for helping you assess progress:** The triangle combines insights from our Routes to Scale framework about what the 'building blocks' are that need to be in place for any system to change, with insights from a wide range of other thinkers and system innovators about the deeper shifts that need to happen for such changes to emerge in the first place, or for them to be sustained (in particular, System Innovation Initiative's Four Keys, and FSG's Six Conditions of System Change). When brought together, these provoke a set of questions about progress that we unpack below.

Using the triangle to evaluate system change progress

The system change evaluation triangle invites a set of questions that evaluators, practitioners and funders can ask to know if their effort is making progress towards genuine system change. If your strategy is focused on establishing 'building blocks', you might need to ask:

- **What is in place to sustain and reinforce this building block, and prevent it from reverting?** For example, is the new funding stream you have unlocked reinforced and protected by a new, widely adopted mental model for why the funding is essential? Is access to a new scaled product or service supported by the relationships needed to embed and sustain it?
- **How has the building block disrupted or shifted the 'raw materials' of the system, and does this shift help or hinder you in reaching your 'North Star'?** For example, has the new supporting legislation you've established caused an informal or formal shift in power to a new stakeholder whose voice is essential to creating meaningful change? Has it facilitated new relationships you need to deliver the outcomes you desire? Has it had unexpected impacts on the 'raw materials'?

If your strategy is focused on disrupting the 'raw materials' of the system, you might need to ask:

- **How are new mental models or beliefs about the system being 'codified'?** What reinforces and amplifies them? Is there consistency in the new language or mental models being used to describe system purpose, and are they being widely adopted?
- **What is changing in practice, as a result of the power shifts you're generating?** Is the power shift helping you move towards your 'North Star' goal? What guarantees or 'locks in' the new sources of power? Does it need supporting legislation or funding to stay in place?
- **What will sustain the new relationships you're creating?** What would happen to these relationships if the people involved changed roles or moved on? Are structures emerging that will help the new norms outlast the people presently shaping them?

Good evaluation as the 'engine' of system change

While system change efforts may make evaluators and evaluative thinkers work a little harder, in our experience, reckoning with the tricky questions thrown up by these challenges is beneficial. It forces us to think more deeply, test our assumptions more rigorously, and engage more empathetically and holistically with the issue at hand. In many ways these challenges are welcome provocations that help us engage more directly with what should surely be the primary question of any evaluation effort: what is the true value of what we're doing? And could we be doing it better?

System change evaluation, when done in a way that accounts for the nature of system work and retains foundational evaluative principles, can become an engine of the work; unearthing new learning, driving adaptation, refining and nuancing strategy, and incorporating new voices and perspectives. The iterative nature of the work, long time scale, and fundamental uncertainty of intervening in complex systems make it especially important to embed structures and mechanisms for deeply understanding what you are doing, and for learning how to do it better. As one system evaluator we spoke to put it: "Developing a clearer sense of what we mean by 'change' and 'results' in efforts to transform systems is a high-stakes challenge. We need to sharpen our thinking about strategy. We need to develop and track indicators of progress so that we can learn from our efforts. We need to communicate our work amongst our allies and those whose support we seek."[33]

This report is shared as a point of departure rather than a final conclusion. We hope to have sketched out a common language to talk about the difficult and complex business of changing systems, and offered some challenges and ideas to inform discussions of and approaches to the practical challenges faced by funders and practitioners, ourselves included. However, we are making just one contribution among many, and do not claim to have landed on 'the answer'. We look forward to ongoing dialogue, constructive challenge, and new practical insights from colleagues, collaborators and experts in the field as we work together to change systems in service of better outcomes for people, society and the planet.

Acknowledgements

We would like to acknowledge the contributions of many thought partners and collaborators who helped shape the thinking and insights in this report. In particular, we would like to thank Zazie Tolmer and our colleagues at the System Innovation Initiative, Jennie Winhall and Charlie Leadbeater whose partnership in 2022 helped us frame the questions to ask and facilitated deep engagement with the field of system change evaluation to begin to answer them. Their sharp insight, evaluation and system change expertise and thoughtful challenge was invaluable in helping us expand and deepen the thinking that forms the basis of this report. We are indebted to the Rockwool Foundation for the financial support that enabled our work together, as well as to current and former members of Rockwool's Evaluation and Interventions teams whose insights, challenge and thoughts were hugely valuable to our exploration – including Helene Bie Lilleør, Anna Folke Larsen, Nina Blöndal and Johannes Björkman.

We would also like to thank the practitioners, funders and evaluators who participated in interviews and surveys, to our project partners in recent years who offered check and challenge to our ideas, shared insights on approaches to system change evaluation, and reflections on where the biggest challenges are, and to our colleagues and peers across Social Finance, whose deep practical expertise across a wide range of sectors and policy areas we have drawn on in many places in this report. We are grateful in particular to Chris Clements, Sara Jones, Louise Savell and Nadine Smith, alongside our wider leadership team, for their roles in developing and refining the work.

Interviewees and collaborators

A full list of interviewees and collaborators is below:*

Peter Batisti, Future of Fish

Leonora Buckland, ESADE

Ingrid Burkett, The Yunus Centre at Griffith University

Mark Cabaj, From Here to There Consulting

Jesper Christiansen, Bikuben Fonden

Yaera Chung, UNDP – Monitoring and Evaluation

Andre Clarke, Comic Relief

Julia Coffman, Center for Evaluation Innovation

Allan Cowie, Chest Heart & Stroke

Anne Dalitz, UNDP - Yemen Country Office

Jess Dart, Clear Horizons

Hugo de Vries, European Federation of Food Science and Technology

Holly Donagh, Paul Hamlyn Foundation

Lillian Flemons, RAND

Jamie Gamble, Imprint Consulting

Jessamy Gould, Treebeard Trust

Penny Hagen, The Lab (Auckland Co-design Lab)

June Holley, Network Weaving Institute

Ed Humpherson, UK Statistics Authority

Liz Jones, Department for Media, Culture and Society

Val Keen, Department of Levelling Up, Housing and Communities

Hannah Lim, Esmee Fairbairn

Amy Ludlow, ShiFT

Cathy Madge, Social Finance

James Magowan, Department for Media, Culture and Society

Joao Martinho, Climate KIC

Kate McKegg, The Developmental Evaluation Institute

Cristina Osoro Cangas, Stone Soup Consulting

Patricia Rogers, Independent Consultant / Better Evaluation

Edward Rowberry, Bristol and Bath Regional Capital

Jakob Schjørring, Bikuben Fonden

Beth Smith, EZC Partners

Anne Bergvith Sørensen, Hjem Til Alle

Ine Steenmans, UCL - Department of Science, Engineering and Public Policy

Søren Vester Haldrup, UNDP Innovation

* NB: while our thinking was influenced, challenged and deepened by many of those on this list, the views expressed in this report are our own and should be not attributed to any on this list, except where explicitly stated.

Selected further reading

Cabaj, M, Evaluating System Change Results: An Inquiry Framework, 2019

CECAN, Choosing Appropriate Evaluation Methods, 2020

CECAN, The Complexity Evaluation Toolkit, 2021

Coffman, J, Stachowiak S, and Lynn, J, Lost Causal: Debunking Myths About Causal Analysis in Philanthropy, 2021

Bolton, E and Davies, J, Changing Lives, Changing Systems, Building Routes to Scale, Social Finance, 2021

Kania, J, Kramer, M and Senge, P, The Water of Systems Change, FSG, 2018

Nesta, Standards of Evidence, 2013

Quinn Patton, M, Developmental Evaluation, 2010

Winhall, J and Leadbeater, C, Building Better Systems: A Green Paper on System Innovation, System Innovation Initiative, 2020

Authors

Jessica Davies

Jessica is an Associate Director based in Social Finance's Impact Incubator, supporting charitable foundations to design and develop systemic solutions to entrenched social problems. She leads our work on methodologies for scaling impact and system change, and developed Social Finance's Routes to Scale Framework. Her work has spanned issues and impact approaches, developing a broad toolkit of strategies over almost a decade's experience of routes to social change, from leading strategy development for the new £1bn Motability Foundation, to convening funder coalitions, scoping and design for the Education Outcomes Fund, investment management and enterprise support in the Care & Wellbeing Fund and rolling out new digital tools to support Children's Social Care.

Jessica joined Social Finance from strategy consulting firm Parthenon-EY (formerly The Parthenon Group), where she worked on corporate strategy and commercial due diligence projects in education, consumer and business services. Jessica also established and led Parthenon's pro-bono practice, with a focus on supporting education/youth charities and social enterprises. Prior to joining Parthenon, Jessica taught Citizenship for two years at a secondary school in West London through Teach First, and spent a year in Beijing learning Mandarin. She holds an MSc in Philosophy & Public Policy from the LSE, and a BA in Philosophy & French from Oxford University.

Madeline Goldie

Madeline is a Manager in Social Finance's Government + Enterprise team. Her work focuses mostly on supporting government and third sector partners to embed practical system change strategies into their work. She was the system change lead for our partnership with the Changing Futures programme – a £64 million programme to transform the way local ecosystems support people experiencing multiple disadvantage and has helped local and regional authorities design and implement radical new commissioning and investment approaches. Madeline is also the project manager for MHEP, a programme supporting social investment in five supported employment services.

Prior to joining Social Finance, Madeline was a fellow with Harvard Kennedy School's Government Performance Lab, where she supported state governments with data-driven reform of the child welfare system, and an advisor for the Australian Department of Prime Minister and Cabinet.

Endnotes

.1 Susan Wolf Ditkoff and Abe Grindle, Audacious Philanthropy, Harvard Business Review, 2017

2 Susan Wolf Ditkoff and Abe Grindle, Audacious Philanthropy, Harvard Business Review, 2017

3 Ashoka, Embracing Complexity; Towards a Shared Understanding of Funding System Change, 2020

4 Cabaj, Mark, Evaluating Systems Change Results, An Inquiry Framework

5 Johnny Morrell, 2021, page 16

6 HM Treasury, The Magenta Book, last updated April 2020

7 See Michael Quinn Patton, and 'Developmental evaluation', (Better Evaluation, https://www.betterevaluation.org/methods-approaches/approaches/developmental-evaluation)

8 MEAM – http://meam.org.uk and Fulfilling Lives – https://www.fulfillinglivesevaluation.org

9 John Kania, Mark Kramer, Peter Senge, The Water of Systems Change, 2018

10 Donella Meadows, Leverage Points: Places to Intervene in a System, 1999

11 System Innovation Initiative, The Four Keys: A Brief Introduction, 2020

12 See, for example, Albertson et al, Payment by Results and Social Impact Bonds, 2018.

13 Nikhil Dugal, Impact Measurement for System Change, Skoll Centre, 2020

14 Better evaluation: What is evaluation? | Better Evaluation

15 Joelle Cook et al, Evaluating Complexity: Propositions for Improving Practice, 2014

16 Department for International Development, Impact Evaluation, 2015; Nesta, Standards of Evidence, 2013; Dartington Service Design Lab, An Integrated Approach to Evidence for Those Working to Improve Outcomes for Children & Young People, 2022.

17 Department for International Development, Impact Evaluation, 2015

18 See the work of Anna Folke Larsen, discussed at the System Innovation Initiative Learning Festival, 2022; Day 3, Session 3: https://www.systeminnovation.org/learning-festival-2022-ressources/wednesday-session3

19 Dartington Service Design Lab, An Integrated Approach to Evidence for Those Working to Improve Outcomes for Children & Young People, 2022, page 9

20 College of Policing, Scared Straight programmes, 2015

21 Zenda Ofir, Evaluation – a Beautiful Practice, 2016

22 Julia Coffman, Oh for the Love of Sticky Notes! The Changing Role of Evaluators Who Work with Foundations, 2016

23 See CECAN, Choosing Appropriate Evaluation Methods, 2020; HM Treasury, The Magenta Book, last updated April 2020; American Evaluation Association, Principles for Effective Use of Systems Thinking in Evaluation, 2018

24 Coffman, Stachowiak and Lynn, Lost Causal: Debunking Myths About Causal Analysis in Philanthropy, 2021

25 Better Evaluation, 2023

26 Emily Gates, speaking at System Innovation Initiative Learning Festival 2022, Day 3, Session 4 https://www.systeminnovation.org/learning-festival-2022-ressources

27 Centre for Homelessness Impact and the International Public Policy Observatory, Homelessness and the pandemic, 2022.

28 See coverage of Álvaro Salas Chaves and Costa Rica's work on community health, covered here: https://www.newyorker.com/magazine/2021/08/30/costa-ricans-live-longer-than-we-do-whats-the-secret

29 System Innovation Initiative, The Four Keys: A Brief Introduction, 2020; John Kania, Mark Kramer, Peter Senge, The Water of Systems Change, 2018

30 System Innovation Initiative, Four Keys https://www.systeminnovation.org/system-innovation-the-four-keys-intro

31 https://www.socialfinance.org.uk/assets/documents/building_routes_to_scale.pdf

32 Skoll, Is it possible to measure system change?, 2019

33 Mark Cabaj

SOCIAL FINANCE

Social Finance Ltd
87 Vauxhall Walk
London
SE11 5HJ
—
info@socialfinance.org.uk
+44 (0) 20 7770 6836

socialfinance.org.uk

Publication date: November 2023

Disclaimer and Terms of Use
This report is not an offering of any Notes for Sale and is provided by Social Finance solely for information purposes. Neither Social Finance nor any of their respective affiliates, directors, officers, employees or agents makes any express or implied representation, warranty or undertaking with respect to this Document, and none of them accepts any responsibility or liability as to its accuracy or completeness. Social Finance has not assumed any responsibility for independent verification of the information contained herein or otherwise made available in connection to the Document. The text in this Document may be reproduced free of charge providing that it is reproduced accurately and not used in a misleading context. The material must be acknowledged as Social Finance copyright and the title of the document be specified.

Social Finance is authorised and regulated by the Financial Conduct Authority FCA No: 497568