A theory-based approach to evaluations intended to inform transitions toward sustainability

Per Mickwitz, Lena Neij, Maria Johansson, Mats Benner and Sofie Sandin
Lund University, Sweden

Abstract
There is an urgent need for radical transformations of unsustainable socio-technical systems, such as food, mobility, and housing. These transformations will not take place without new policies and research. In order to achieve these transitions, learning must be a central feature based on thorough evaluations of the actions taken. Evaluations have been conducted and studied for decades, but traditional evaluation approaches have largely been developed to produce knowledge for incremental changes, not for radical transformations. This article develops a framework for interdisciplinary evaluations targeting transformative changes toward a more sustainable society. The framework combines evaluation theory and practice with transition theory, sociology of science, policy analyses, and environmental psychology. While the primary purpose of the framework is to help design evaluations that would better enhance learning for transitions, it can also be used for systematic meta-evaluations of past evaluations.

Keywords
theory-based evaluation, transition theory, sustainability, meta-evaluation

Introduction
While the development of modern societies has provided enormous benefits for many, it has also resulted in major environmental, social, and economic challenges. Rapid technological change and increased general wealth have as side effects produced significant losses of natural capital, accelerated climate change, but also large income differences. To switch to
a more sustainable development path, a transition in society and in systems of consumption and production is urgent. In 2015, the United Nation’s Sustainable Development Goals (SDG) were adopted and are now to be realized through Agenda 2030 worldwide. This transition will require not only the introduction of new technologies but also changes in the entire socioeconomic structure involving institutions, social network, individual behavior, and acquirement of knowledge.

To realize the transition, political strategies and policy instruments are required. Policies are needed to overcome market failures, provide new knowledge, and accelerate and direct changes in sociotechnical systems. In order to be effective, such policy interventions need to be evaluated to facilitate learning based on the actions taken. Whereas traditional policy interventions and concomitant evaluation approaches have focused on incremental changes, transformative changes are more complex and will require new approaches in policy design as well as in policy evaluation.

Evaluations have been conducted—and studied—for decades. The link to and use of social science in evaluations are too rare, and when used, the evaluation approaches are frequently only linked to just single disciplines. In order to capture the complex changes required in the sociotechnical systems and to create learning necessary to understand transformative changes, new types of evaluation frameworks need to be developed. These frameworks have to go beyond traditional disciplinary methodologies and rely on interdisciplinary approaches.

Our aim is to build on the theory and practice of evaluation, and to develop this further by demonstrating the importance of combining evaluation theory and practice with other relevant theories. The specific choice of which additional theories to use will depend on the subject of the evaluation. The interdisciplinary evaluations framework can be used in evaluating new policy interventions as well as to assess past evaluations.

Many sustainability crises are caused by the present unsustainable consumption and production systems (e.g. food, energy, mobility, and housing), which have depleted natural resources and increased the production of waste and emissions (European Environment Agency (EEA), 2019). These systems emerged through technological innovations and the sociotechnical systems built around them. More sustainable systems will require not only new technologies but also renewed institutions and changed individual behavior. The interdisciplinary evaluation approaches and framework developed and presented build on evaluation theory, combined with transition theory, policy analyses, sociology of science, and environmental psychology (Figure 1). These are four research approaches and theories we find central for evaluations providing knowledge for transitions toward sustainability. We recognize that many more research approaches could be included, and that this first iteration of the evaluation framework should be extended and improved. Interdisciplinary work is challenging, but the possibilities for success can be hugely improved if there is a common goal and a coherent joint framework for knowledge integration. The common goal is to produce evaluation-based knowledge for transitions toward sustainability. The joint framework for integrating disciplinary theories, approaches, and insights is built on the foundations of evaluation theory.

Donaldson and Lipsey (2006) argue that evaluations may use three types of theory: evaluation theory, social science theory, and program theory. Adopting this approach, we start out from evaluation theory, building on scholars such as Shadish et al. (1995), Vedung (2010), and Alkin (2013), complementing the framework with additional theory-based elements essential for the evaluation of a transition toward more sustainable sociotechnical system. In doing so,
we introduce social and behavioral theory, and as mentioned we use input from transition theory, policy analyses, sociology of science, and environmental psychology. While these theories influence many aspects of evaluation (focus, methods, criteria use, etc.), only when they are also utilized in program theories can the approach be labeled “theory based.”

A theory-based framework for evaluations for transitions

Evaluation theory has developed over the years and is based on an exploding practice colonizing almost every field of human activities today. In parallel with—and in connection to—the expansion of the practice of evaluation, a rich body of evaluation theory has emerged. By combining the question-based approach to evaluation practice by Vedung (2009) with the three main branches—methods, valuing, and use—of evaluation theory (Alkin, 2013; Shadish et al., 1995) a generic evaluation framework has been developed (Table 1). This framework is then utilized to explore where and how theory-based knowledge should be considered in order to design evaluations for transitions. The presented framework consists of five overarching aspects (A–E in Table 1) with a number of specific questions for each aspect. These main aspects and specific questions can be used to initiate and design evaluations in a conscious way, to interpret the outcome and compare it with previous results, as well as to assess the overall quality of evaluations already made.

Evaluation context

The proposed framework recognizes that the outcome will always be dependent on the specific content of the particular evaluation. The first aspect included in the assessment framework is thus the context of the evaluation (A in Table 1).

The specific questions related to the evaluation context have largely been inspired by the 11 Questions Approach to Evaluation by Evert Vedung (2009). The specific questions are included in Table 1 (A). The question of the purpose of the evaluation covers such aspects...
### Table 1. A framework for assessing evaluations from a transition perspective.

| Main aspect | Specific questions |
|-------------|--------------------|
| **The context of the evaluation (A)** | When is the evaluation undertaken in relation to past and future decisions? (A.1.)  
What is the main purpose of the evaluation? (A.2.)  
What is the relationship between the evaluation and other (past and future) evaluations and the overall evaluation strategy? (A.3.)  
Who is commissioning the evaluation? (A.4.)  
Who is doing the evaluation? (A.5.) |
| **The focus of the evaluation (B)** | What is the evaluand and how is it described? (B.1.)  
Which other explanatory factors and important features of the context are specified? (B.2.)  
What is the temporal limitation and focus of the evaluation? (B.3.)  
Is the focus on intended effect or are side effects (intended and unintended) and rebound effects also considered? (B.4.)  
Is there any focus based on transition theory (e.g. experiments, niches, path dependency, or vision)? (B.5.) |
| **The design, methods, and data used to assess impacts (C)** | What are the fundamentals of the evaluation design? (C.1.)  
How is theory used? (C.2.)  
What empirical material is collected and analyzed? (C.3.)  
How have impacts been assessed? (C.4.)  
Have methods and material for assessing the distribution of impacts and costs among different groups been applied? (C.5.)  
How is the counterfactual constructed (is the full policy mix and all governance levels taken into account)? (C.5.)  
How is it determined that side effects, at least partly, are due to the evaluand? (C.6.)  
If triangulation (of theories, methods, and data) is used how has the synthesis been produced? (C.7.) |
| **The criteria used for valuing (D)** | How and by whom have the criteria been decided? (The organization commissioning the evaluation, the evaluator(s), by stakeholders, general evaluation policy, . . .) (D.1)  
Which value criteria were used to judge the intervention? For example, relevance, effectiveness, efficiency, flexibility, predictability, persistence, acceptability, transparency, and equity (single- or multi-criteria approach). (D.2)  
Do the value criteria reflect the interests of different groups, including poorly organized and powerless? (D.3)  
Do the criteria used promote reflexivity and challenge established goals and the framing that policy is based on? (D.4)  
Is reflexivity part of the value judgment the conclusions are based on? (D.5) |
| **The approaches to facilitate use (E)** | Have key stakeholders been identified and involved in the evaluation process? (E.1.)  
Have there been any specific efforts to engage different groups, including those that are not well organized and powerless? Do these efforts target also other levels of governance than the one for which the evaluation has been commissioned? (E.2.)  
What has been the time frame for the use of the results? (E.3.)  
What particular activities have been undertaken to facilitate use? (E.4.)  
Have there been efforts to promote use beyond “intended use by intended users” by making the process open and transparent or by making the evaluation results/report freely available and easy to obtain? (E.5) |
as to what degree the evaluation is undertaken for some predefined decisions and to what extent it is made to create a general knowledge base. Some of these contextual issues set boundaries for other aspects, for example, when an evaluation is undertaken, or its main purpose may limit the data that can be used or may influence the approaches taken to facilitate use. Other contextual factors will mainly be used when analyzing and interpreting the material. If certain methods are frequently used, can this be due to many evaluations being commissioned or conducted by the same actors? The understanding of the evaluation context is also significant in relation to how the results will be perceived by, for example, different stakeholders, especially if there are conflicting interests around. Stakeholders’ and the public’s view on the purpose of the evaluation and trust in the organization behind the evaluation are critical for the legitimacy of the results and ultimately the behavioral change that can be expected.

Although evaluators might sometimes introduce theories, methods, and expertise that surprise those commissioning an evaluation, the general purpose of an evaluation (A.2.) can largely limit or enhance the use of insights from theory in the other aspects of the evaluation. In other words, if large-scale transitions, research, and policy-induced development and behavioral change are not considered at all when the purpose is determined, theoretical insights about these are not very likely to be used. For example, transition theory would suggest evaluations with the purpose of assessing whether policies contribute to large-scale system change, but also evaluations related to whether policies contribute to inertia and support dominant unsustainable systems.

**Evaluation focus**

The second aspect in the framework, that is, the focus of the evaluation (B in Table 1), is based on the importance given to framing in policy analysis. The first question (B.1. in Table 1) is where evaluand refers to the focus of an evaluation in the evaluation jargon (Scriven, 1991: 139). Evaluations may be focused on a project, program, or policy or even on a mix of them. Is the evaluand described in terms of policy instruments, recognizing their main features (regulation, economic, information, and innovation push/pull)? Is the full policy mix, including policies introduced at other levels of governance, recognized and considered when the evaluation is focused? An evaluation could focus on one process only or the merit, worth, and value of the process in generating outputs, for example, energy-efficiency certificates. More frequently, however, the interest is related to the wider outcomes and impacts of projects, programs, and policies. The description of the evaluand also covers issues such as the intervention target group socio-demographic, the actors producing the intervention, the main logic of the intervention, and how it was introduced. It is also important to understand the motivations of the target group individuals, such as their knowledge, values, norms, and attitudes. In relation to the introduction for an individual’s perspective, the key issues are the concerned people’s involvement and participation in the process.

Policies tend to have side effects. An important question is thus to what degree the evaluation is focused on revealing side effects, including unanticipated effects not previously even known of. This is covered by the specific question in B.4. of Table 1. Here, it is worthwhile to be aware of both the intended and unintended effects as well as the spillover effects including rebound. Transition theory has clearly established the importance of certain features for radical transformations of sociotechnical systems. The question in B.5. of Table 1 covers these.
Evaluation design, methods, and data

A key aspect of evaluation theory concerns the methods utilized to generate results. The third aspect of the framework is therefore the design, methods, and data used to assess impacts (C in Table 1). Question C.1. refers to both the use of social science theory and program theory, and the use of social science theory in program theories. Of particular interest here is whether and how transition theory, sociology of science, policy analyses, and environmental psychology theories will be used. These theories can be used at a general level or for particular aspects of the program theories. For example, what is the theoretical foundation for the choice and operationalization of behavioral antecedents or behavioral intentions considered. The second question (C.2.) relates to whether the evaluation is based on an experimental or quasi-experimental design, case studies, and so on. This question is further elaborated in question C.4. The assessment method determines the empirical data collected and used, but at the same time the data that is available or can be collected limits the methods that can be used. One cannot do sophisticated time-series analyses, with very few observations.

The full scope of documents, surveys, measurements, statistics, interviews, focus groups, and participatory observation is covered in question C.3. By asking “How has the impact been assessed? (C.4.),” the general evaluation design is made more specific, addressing how the attribution of the intervention is separated from that of other factors. This is also closely related to the question posed in C.5: what would have happened without the specific intervention? The final question in this section concerns synthesizing multiple data, theories, methods, and perspectives.

The construction of counterfactuals may also differ. Counterfactuals can be baselines that are either calculated or extrapolated based on certain data and assumptions or reflecting the outcomes before a certain policy instrument came into effect. They can also be in the form of reference groups or constructed from answers in surveys and interviews. One challenge thus lies in the assessment of the scale presented by the counterfactual and its reliability.

The criteria for valuing

Evaluation concerns value judgment and a branch of evaluation theories are centered on valuing. The criteria for these judgments are important, which is the focus of the fourth aspect of the approach (D in Table 1). The first question gives a base to compare single and multicriteria evaluations. Not only are the criteria used important, but so are the legitimacy and interests reflected in the criteria. The questions “To which degree do the value criteria reflect the interests of different groups, including poorly organized and powerless?” and “How and by whom have the criteria been decided?” are included to cover these aspects. These are questions that need to be carefully considered in relation to the context of the evaluation as this may largely differ between settings and situations. As one key function of evaluation is to provide input for learning, question D.2. is explicitly included to examine how the potential of evaluations to enhance double loop learning (Argyris, 1999) has been used.

The criteria used in an evaluation largely determine the merit of the evaluand. There are various sets of criteria available, either general or related to democracy or economy, for example (Mickwitz, 2003), or more specific related to expectations on individual behavior. A common general criteria, however, evolve around some sort of calculated impact (in, e.g. saved kWh or reduced amount of CO₂-emissions), as well as the cost-effectiveness of each saved
unit, but also a desire to, for example, facilitate, inform, and inspire receivers toward a path that will generate a change in the energy consumption patterns may be expressed. With aims such as carbon neutrality frequently being adopted (EEA, 2019), the relevance or even possibility of a technology or social practice to be part of a carbon-neutral society becomes an important criterion. Simultaneously assessing the cost-effectiveness of reducing emission becomes less relevant.

It is also vital to not be blinded by the program goals and impacts, but to also use criteria that can promote reflexivity on whether the policy in question is relevant to the problem, and whether established goals are in line with norms and visions. A program intended to, for example, increase the amount of district heating may boast of brilliant results in an evaluation, but is it directed toward a sustainable path or is it instead locking out other available solutions?

Facilitating use of an evaluation

The third branch of evaluation theory and an essential feature of evaluations is based on use; thus, the fifth and final aspect of the framework is the approaches to facilitate use (E in Table 1). Based on key evaluation theories about use, specific questions have been included. Since use—especially in political contexts—can happen at different timescales (Valovirta, 2002), a query on the time frame is included.

This part of the framework steps away from what is mostly present in evaluation reports and seeks in a sense instead to determine what is not there. In the task of determining whether key stakeholders have been identified and involved, it has to be defined who key stakeholders are and to what extent they should be involved in order to provide a meaningful contribution to the result. One challenge is thus to map the intervention and identify the actors at the different stages, and then to determine whether they have been involved in the evaluation sufficiently. This raises the issue of what is to be deemed sufficient involvement in terms of direct (e.g. discussion, interpretation) or indirect (e.g. answering a survey) participation and how the introduction of the intervention has been framed to these groups.

The second part of the aspect of facilitating use evolves around enabling use of the results of the evaluation, and the table includes a question to investigate the link between evaluations and informed democratic deliberations (E.5. in Table 1).

The time frame for the use and accessibility of the results and the transparency of the process are key concepts to opening up to a broader audience. For example, an evaluation of subsidies for biogas-based heating will naturally be of interest to the government that has financed the subsidies, but may also spark initiatives among external actors for further development of alternative heating sources more suitable for other geographical areas or demographic groups. Thus, efforts to promote use beyond Patton’s (2008) “intended uses by intended users” are of major interest from a transition perspective in particular.

Social and behavioral science theories as a knowledge base for transitions

There has been a general plea for more social science theory in evaluations (Donaldson and Lipsey, 2006). The need for social science theory is, however, especially vital if the aim of evaluation is to produce knowledge for transitions toward sustainability. The reason is that if we want knowledge of system-level radical change, empirical data and models estimated
Based on past relationships will not be sufficient. The only way to produce meaningful insights on radical change is by utilizing theories of the essential dynamic processes. In order to generate knowledge through evaluations on how policies might contribute to transforming large sociotechnical systems (food, energy, mobility, housing, etc.), we need a theoretical base from transition theory.

Fast and directional transitions require policies; we therefore also need policy analyses as a basis for understanding the dynamics and key features of the political processes through which they are formed, implemented, and evaluated. Technologies are key aspects of sociotechnical systems. Through innovation, technologies develop rapidly and science is often an important factor. Science also influences social practice and policymaking. This is why the sociology of science is an important theoretical foundation for evaluations for transitions.

Human behavior is a main factor in the sustainability crisis; consequently, policies that impact human behavior must be sought (Gardner and Stern, 1996). This implies that behavior should constitute a key factor in any evaluation of initiatives toward transition to sustainability (Steg et al., 2015). Yet, the consequences of individual behavior are often overlooked in the discussion (Stern et al., 2013). Behavior can be studied from many different perspectives: sociology, economics, and psychology. We have here chosen to focus on psychology, mainly because it is a perspective that has been sparsely utilized in relation to sustainability transitions but also because it is particularly relevant for energy efficiency in buildings, which was the empirical case our framework first was utilized for (Neij et al., 2021; Sandin et al., 2019).

While there are strong arguments for utilizing transition theory, policy analyses, sociology of science, and environmental psychology in evaluations for transitions, there are also other fields that can contribute, such as social practice theory, planning theory, or ecology. The use of the theories from the four approaches chosen demonstrates how also theories from other fields can be integrated in a theory-based evaluation framework for evaluations for transitions.

Insights from transition theory, important for enhancing radical system transformations

Several theoretical approaches focusing on changes in sociotechnical systems have been developed since the 1970s. Among these are, the theory of large technical systems (see, for example, Hughes, 1983), the theory of social construction of technical systems (see, for example, Bijker et al., 1987), the theory of actor network (see, for example, Callon, 1987), the theory on innovation systems (see, for example, Bergek et al., 2008; Carlsson, 1995), and the theory of a multilevel perspective (Geels, 2002, 2004). These theories all describe the technology in a social context—each theory with a different focus and approach. Based on these theoretical approaches, transition-based theories were introduced in the 1990s focusing on major (radical) technical changes and the assessment of governance approaches toward a more sustainable society. In all, transition studies capture many different theoretical approaches, with similarities but also clear differences (e.g. Turnheim et al., 2015; van den Bergh et al., 2011).

Here, we highlight three concepts of these theories that we find central—visioning, experimenting, and learning. We also highlight the need for a system approach, a multi-actor perspective, and the consideration of scale.

Visioning, to picture new innovations and emerging systems, is emphasized within transition theory (Berkhout et al., 2004; Rotmans et al., 2001). It represents the long-term aspirations for a transition and aids in providing arenas for experimentation. The time aspect
connected to visions is of importance since the establishment of new technology or reconfiguration of sociotechnical systems takes time, whereupon the phase of the change process needs to be acknowledged.

Many of the studies of transitions have established that during early phases there are often many different partly competing technologies and practices of which most fail, but some develop and transform into essential elements of new sociotechnical systems or arenas (Jørgensen, 2012; Smith et al., 2005). The important features of these dynamics are first that there are enough experiments (Kemp, 1994; Loorbach, 2007; Schot and Geels, 2008) and that they are different enough, that is, there has to be variation. Second there needs to be selection, that is, not every experiment will continue. However, how and at what stage, and based on what criteria the selection takes place, is crucial. While a lot of experimentation is taking place, newer research stresses the importance of embedding them through institutionalization, circulation, scaling up, and replication, in order to enhance their role for transitions (Turnheim et al., 2018).

A novel product, service, technology, or practice is seldom immediately developed enough to be able to successfully compete against already established ones. Specific spaces—niches—are required where further development can take place (Geels, 2004; Raven, 2007). These spaces for change are, for example, described as the microlevel in a hierarchical system where ideas can grow in protected isolation (Raven et al., 2012) or as the ongoing activities within an existing arena, pending for a momentum to build up enough to overthrow the prevailing construction (Jørgensen, 2012).

Furthermore, there is a large emphasis on learning (and reflexivity) by many scholars in the transition field (Geels and Raven, 2006; Kemp, 1994; Rotmans et al., 2001). Learning can be achieved through experimentation: by trying, evaluating, and revising. Learning is also essential for accelerating change, by diffusing and institutionalizing successes and further improving them. In order to fuel learning processes within the public policy area, evaluation of policy instruments is key. The scope of evaluations may vary, and even though evaluations of policies, programs, and projects are undertaken primarily to achieve incremental improvements or to legitimize the status quo, they might yet contribute to radical transitions in the buildup of a knowledgebase.

The incorporation of a holistic system approach is prevalent within transition theories, where linkages between different levels (Geels, 2002) or the interplay between different actors, institutions, and technological factors are central. Transformations in a system may come about through, for example, innovations stemming from experimentation and niches, or from institutional changes. Crises or ruptures in a beforehand stable constitution can, however, open up spaces for emerging alternatives and changes (Jørgensen, 2012). System transformations are, therefore, often largely affected by general forces, often external to the particular system. For example, radical changes in the price of oil have been essential for both the development of renewable technologies and energy-efficiency solutions. Other examples of these general factors are changing values and norms, new scientific discoveries, general policies, for example, trade agreements, or shocks such as the nuclear accident in Chernobyl. Within the multilevel perspective, these factors are often referred to as landscape change, that is, factors occurring at a level exogenous to any particular sociotechnical system (Geels, 2002). Thus, it is widely recognized that transitions cannot be planned and controlled, but the actors involved can learn and influence each other and the pathway of the transition.
The existing sociotechnical systems are maintained by forces upholding the prevailing con-
stitutions, and for changes to come about these constructions need to be challenged. The realiza-
tion of a transition thus requires a destabilization of the path dependences and lock-ins of
the present unsustainable systems (Kivimaa and Kern, 2016).

A holistic system approach calls for a multi-actor perspective (Raven et al., 2012; Rip and
Kemp, 1998; Schot et al., 1994; Smith et al., 2005). A transformation brings changes that
affect a multitude of actors, but actors may likewise influence the change process, by either
aiding or counteracting it. Such dynamics may be captured only if the approach allows for
multiple actors from different corners of a sociotechnical system to be acknowledged.

A final key aspect of transitions is that they depend on developments at different scales
(Geels, 2002; Raven et al., 2012). Technology and knowledge are largely global, but at the
same time they are developed and used locally, and factors such as governance, markets, cul-
tures, and networks may exist both on local and global scales, thus creating a vast space of
innovation (Raven et al., 2012: 69). Scale is, however, not only a geographical concept, it also
concerns specific innovations of products, services, technologies, practices, and the change of
entire systems in the temporal scale, linking the immediate development with the long-term—
often decade long—processes (Turnheim et al., 2015). The act of evaluating with regard to
transitions thus requires the integration of different scales, from single technologies to sys-
tems, combining local efforts with global developments, and being sensitive to different tem-
poral developments where some changes occur rapidly and others take many decades.

The sociology of science and its relevance to smart evaluations

Transitions hinge upon knowledge production. It enables variation and experimentation,
affords reliability, underpins assessments, and trains and spreads qualified scientists, engi-
neers, and analysts. But how can its impact on transition processes be judged? The relation-
ship between research and transitions has been conceptualized differently over time. The
classical conception of how research brings about change in society is the linear model out-
lined by Bush (1945). It asserted that basic research—if properly validated—will result in
pieces of information that can be communicated and later on translated into applied research,
which adapts that general information for specific settings. Applied research subsequently,
like in a relay race, enables development work which refines the applied communication into
blueprints for production, legislation, and other practical interventions. Finally, practices of
various sorts scale up these blueprints into real artifacts.

The Bush thesis emerged at about the same time as the sociology of science literature began
its meteoric rise under the guidance of Columbia sociologist Robert K. Merton (1973). For
Merton and his followers, the scientific system was largely self-organized around a set of
norms and social hierarchies, which asserted its efficiency and safeguarded its autonomy from
political steering. From such a perspective, basic science could be governed and evaluated on
its own terms while still contributing to social change—through the linear process outlined
above. If only basic research was allowed to flourish, the entire society would benefit.

The Bush/Merton axis remains an important foundation for both research policies and eval-
uations. The European Research Council (ERC) and a string of other research policy initia-
tives are premised on the notion that society will benefit from high quality research, as defined
by colleagues (Kaiser, 2017). The role of such research, labeled frontier research, high risk-
high gain, and so on, is in the words of ERC’s former president, Helga Nowotny:
not only to explore the yet unknown, but also to be aware that the frontier must be linked to its hinterland. If new discoveries are made at the frontier, innovation turns them into new products and processes. (Streicher, 2006)

Transitions are therefore, from this perspective, critically dependent on cognitive variation, as brought about by truly innovative and novel inquiries.

The Bush/Mertonian tradition lingers on and has a profound—if implicit—impact on how many research evaluations are conducted and how the contributions of science are viewed more generally. As the view is that science evolves within intra-systemic mechanisms, such as communication and reward models (which reflect the relative importance of a contribution to the “scientific community’s” reproduction) (Hagstrom, 1965), the relative importance—or quality—of a research effort can be understood by either qualitative assessments, such as peer review (collegial analysis), or by quantitative indicators (bibliometrics), while the social mechanisms behind its impact can be modeled, controlled, and understood (Moed, 2006). The main criterion for evaluation is whether a specific piece of research is accepted (cited and/or esteemed) by the scientific community to which the evaluand adheres. Indicators capture the essence of a publication’s contribution to the scientific community and can be used to gauge that. The rationalist sociological tradition thus seeks to elucidate the mechanism by which the research system functions, in terms of resource allocation, promotion (and demotion), and rewards, in order to reproduce itself. The most desired outcome of that is of course the production of truly innovative and path-breaking research, which not only shapes the intellectual trajectories of the respective fields but also opens up new avenues in society (Zuckerman, 1977).

The Bush perspective came under heavy criticism in the 1960s, where first a reversed linear model was afforded which claimed that most inventions (innovations, transitions) have little or no relation to basic research but rather emerge out of economic necessities and demand in society (Schmookler, 1962). Later, increasingly sophisticated versions of coupled models have tried to capture the relationship between demand, knowledge production, social interests, and other intervening factors. Many of them—or their descendants—have already been discussed here, but there are parallel debates in science studies which try to capture the dynamic relationship between research and social change—and which therefore highlight elements of how research and society co-evolve (Gibbons et al., 1994). Such perspectives afford also a more fluid reading of how science evolves—in a combination of intrinsic knowledge development and communication within society. The main aim for evaluation is therefore to assess the mechanisms of this co-evolution—interaction and engagement with stakeholders, dynamics in problem formulation, reflexivity in assessments, structures of agency, and openness for variation and change in multi-actor and multi-purpose constellations. The recent heavy engagement with the impact of research (HEFCE, 2016) clearly belongs to this policy paradigm, with its emphasis on the social embeddedness of research. It also lends itself to a linear conception as it assumes that research can have a direct and unmediated relation to societal utility (Martin, 2016).

A congenial evaluation approach would highlight how institutional settings form and direct scientific work, for instance, how models of funding and organizing research shape scientific practices (and, reversely, how practices shape institutional arrangements) (Pestre, 2003). It studies scientific work as the interplay between purposeful actors and constraining institutions—and does so by emphasizing such aspects as how problem choice is made (which areas are funded and why), how problem choice is operationalized (through bodies controlled by
scientists, or by bureaucracies, or by stakeholders, and so on), how funding instruments are devised (centers, programs, projects), methods of priority-setting (by whom, in which arenas, and with what focus), the governance of scientific practice—collaborative versus individual activities, methods of appraisal (qualitative or quantitative)—all in all, a focus on the power and relations of knowledge production, and how they altogether affect the research process and its articulation with broader goals such as transition (Frickel and Moor, 2006). Thereby, it showcases both rational and social elements in the evolution of scientific work and how transition forms one framework for knowledge production—and vice versa: that knowledge production with transformational aims impacts transition processes (e.g. policy, behavior).

**Insights from policy analysis in relation to policy interventions**

Transitions toward more sustainable systems have for many years been supported by public policy and a number of policy instruments. In general, public policy, that is, what the public sector aims to achieve, the means or instruments by which the aims are to be reached, and the implementation of the instruments, can largely influence societal development. The decrease of traditional point source water and air pollution in Europe can largely be attributed to public policies (EEA, 2019), but so can the increased inequality in the United States (Hacker and Pierson, 2010).

There are many ways to categorize policies. The most frequently used is to distinguish policies based on the underlying logic of the policy instrument. This results in a distinction between regulation, which determines what one is forbidden to do or allowed to do if a permit has been obtained; economic instruments, which remove resources through taxes or fees or give additional means through subsidies or provision of services; finally, action might be influenced through information (education, information campaigns, labeling, etc.) (Vedung, 1998). Policies can, however, also be classified based on their aims, that is, policies that aim to reduce greenhouse gases are considered climate polices and policies that intend to produce innovations are innovation policies or based on the actual effects of the policies, in which case a speed limit that cuts fuel consumption can be seen as a climate or energy-efficiency policy.

In relation to innovations and system transformations, policies are often divided into push and pull policies (Dosi, 1982; Peters et al., 2012). Push policies support innovations and new technologies through research, support for R&D, or investments, that is, action that increases the supply, whereas pull policies aim at creating demand for innovations by, for example, public procurement, regulation, or taxation. Nowadays, there is an increasing emphasis on the need for policy mixes in order to enhance system transformations (Kern et al., 2017; Rogge and Reichardt, 2016). Policy mixes refer to not only a combination of push and pull instruments but also to instruments that address different actors in the system as well as combinations of regulations, economic instruments, and information. The policy mix concept is, however, broader than just the instrument mix; it also covers the policy process through which these emerge and develop and some key characteristics of the instruments and the process, such as consistency, coherence, credibility, and comprehensiveness (Rogge and Reichardt, 2016).

Although policies may have huge impacts, this is not always the case. Lack of impact may be due to policies being based on wrong assumptions (Hoogerwerf, 1990), problems related to implementation (Pressman and Wildavsky, 1984), or dynamic contexts, that is, policies are implemented in circumstances different from those for which they were planned. These three
possible sources of policy failure are explicit in the program theory framework by Chen (2005), where the change model incorporates policy assumption, the action model the implementation, and these two interlinked models play out in a dynamic context.

Policies do not just have the desired impacts or fail. A key feature of policies is that they tend to have side effects, some of which are anticipated while others are un-anticipated, some are in the same area as the main aim of the policy, some are affecting the target group, other side effects concern other groups in the society, and some may be judged as beneficial while others can be seen as detrimental (Vedung, 1997). Popper (2003: 105) has argued that

the main task of the social sciences . . . is the task of analyzing the unintended social repercussions of intentional human actions.

A particular side effect especially important in relation to energy use and energy efficiency, but also more generally to resource use and sustainability is the rebound effect. The rebound effect refers to the fact that when energy efficiency (or resource efficiency) is improved, money is also saved due to the reduced cost of energy. The saved money is either used to expand the same activity (e.g. driving) or it is used on something else, which directly or indirectly uses energy (or resources). Due to the rebound effect, the energy saving becomes smaller than first assumed (Gillingham et al., 2015).

An important aspect of policymaking is setting the agenda and forming policies; these processes (as well as also other social actions such as evaluation) are largely affected by framing. Hajer and Laws (2006) state that a “frame is an account of ordering that makes sense in the domain of policy and that describes the move from diffuse worries to actionable beliefs.”

Traditionally, public policies were largely formed by the national state although implementation often took place regionally or locally. The situation has, however, changed with a lot of policies being formed internationally, for example, in the United Nations or within the mandate of international agreements, and extensive policy formulation occurring in the European Union. This type of multilayered policy is often referred to as multilevel governance (Bache and Flinders, 2004). Recently, the concept of polycentric governance gained even more popularity when describing and analyzing policies formed and implemented by different actors at different locations (Schoenefeld and Jordan, 2017). In addition to capturing the increased interdependence of policymakers at different levels, the multilevel governance and polycentric governance concepts stress the interdependence of governmental and nongovernmental actors at different levels. An important feature of these types of governance is that the interactions are not only unilateral from the top down. For example, the local action taken by municipalities and cities to mitigate climate change has showed possibilities and put pressure on states and global negotiations.

Policies are always political. One aspect of this is that powerful interests with strong resources to lobby and influence policies can often influence the aims and the specific details of policies. An example is that environmental taxes in almost all countries have exemptions for big export industries (Hogg et al., 2016). In most political settings, no actor alone is powerful enough to make decisions, and policymaking is thus based on coalitions (Sabatier and Jenkins-Smith, 1993). The political coalitions might be more or less stable. They might be wide or issue-specific and they might, besides politicians, also include interest groups, authorities, businesses, and experts. The compositions and dynamics of policy coalitions depend on context, policy style, and the institutional setting. In the analysis of international environmental policymaking, epistemic communities have been used to analyze how networks of experts
with recognized knowledge and some shared beliefs, on casual relationships and relevant knowledge, are essential to enable political action (Haas, 1992). In democracies, politicians gain legitimacy through the way they are elected. Other participants in the policy process (e.g. authorities and public agencies involved in policy preparation and implementation) do not have the same source of legitimacy. This stresses the importance of process and output legitimacy (Skogstad, 2011). Evaluation is an activity that can largely influence the legitimacy of policies as well as actors. The influence on legitimacy may be intended or unintended consequences of the evaluation. At the same time also, the evaluation is largely dependent on its own process and output legitimacy. If the legitimacy of an actor or a policy is largely lost, due to, for example, a scandal or following a gradual decrease in its social acceptance, this may change the political power balance and open new windows of opportunities for policymaking as well as evaluations.

**Environmental psychology to understand antecedents of environmental behavior**

The importance of considering people’s behavior in transitions toward sustainability is stressed in several disciplines, including economics, psychology, and sociology and from diverse theoretical perspectives on human action. The common agenda is to understand the circumstances that motivate action in a more sustainable direction (Moezzi and Janda, 2014). Evaluations providing knowledge for sustainability transitions would benefit from pinpointing under what conditions sustainable actions are supported and when they are hindered, and what factors may trigger transition and maintain such actions. Humans as individuals and their daily life choices will collectively have huge effects on the overall outcomes of any policies implemented. Here, we have chosen to depart from an environmental psychology approach to discuss how sustainable behavior and its antecedents could be thoroughly analyzed in evaluations. Reality is complex and may benefit from theoretical as well as methodological triangulation in evaluations of sustainable behavior (Chatterton and Wilson, 2013; Johansson et al., 2020). Here, the environmental psychological perspective is used as an example, but does not mean that other perspectives on behavior should be disqualified in evaluations of sustainable behavior.

Environmental psychology is engaged with the interaction between humans and their physical settings. It draws on a wide set of psychological theories but always considers the dual relation between how individuals change the environment, and their behavior and experiences are changed by the environment (Gifford, 2014). Such transactions concern both the physical and social characteristics of a setting. Environmental psychology is multidisciplinary and makes use both of theories that consider people’s behavior as a response to new technology per se and theories that regard behavior as embedded in a sociocultural context (Lenoir-Improta et al., 2017). The focus is, however, on the psychological processes antecedent to the behavior, taking into account intra-individual as well as external contextual factors. Thereby, environmental psychology research offers a nuanced description of behavior from the individual user’s point of view. This approach allows for precision of the evaluation of behavior by proposing theoretical frameworks and concepts to understand the complexity of people’s behavior as individuals, the motivations for their behavior in terms of psychological antecedents of the behavior, and the psychological processes of behavioral change in response to policy interventions.
Sustainable behavior can be understood as behavioral intentions, observable (overt) behavior, as well as the broader implications of performed behavior for quality of life and well-being. Behavioral intentions concern what people say they intend to do or would be willing to do, but does not necessarily translate into overt behavior, for example, there is likely to be a gap. Behavioral intentions are sometimes operationalized as acceptance of or willingness to pay for sustainable technologies (Huijts et al., 2012; Steg et al., 2015). From a psychological point of view, overt sustainable behavior can be differentiated between behaviors that involve the adoption of more sustainable equipment and behaviors that involve the recurrent use of equipment (Gardner and Stern, 1996). The former is referred to as efficiency behavior and typically implies a single action, whereas the latter is termed curtailment behavior and implies behavior changes on a frequent basis (Schuitema and Bergstad Jakobsson, 2013).

There is an ongoing discussion about how context-specific sustainable behavior is over different domains. One way to understand this is to consider spillover to other behavior. Some studies suggest that one type of sustainable behavior may inhibit other such behavior (referred to as negative spillover). Other studies suggest that there is a positive spillover effect where the new sustainable behavior also initiates other related behavior (Truelove et al., 2014). Many daily life behaviors are habitual, for example, carried out automatically in a certain context without much cognitive elaboration. Habits facilitate daily life but constitute a challenge in behavioral change as the individual must first be made aware of the behavior. Changing energy use behavior such as changing heating practice may also have other side effects on daily life. Perceived quality of life is a multidimensional construct in environmental psychology, defined as the extent to which important values and needs of people are fulfilled in terms of social relations, pleasure, work, health, privacy, money, and status, but also to different aspects of the physical environment such as access to clean air and soil, and the presence of plants and animals, important to well-being and health (Moser, 2009). New overt behavior may also have broader implications to perceived quality of life that may hinder or support maintenance of the behavior.

People’s engagement in sustainable behavior is not just limited to instrumental factors, for example, costs and benefits in terms of price, time, and comfort. It can also be motivated by affective factors and by social costs and benefits. Such antecedents of behavior refer, for example, to people’s perceptions, emotions, values, norms, and attitude (Gifford and Nilsson, 2014; Steg et al., 2015). Moreover, the role of the social context and the physical environment factors has been stressed such as the role of place attachment (Lenoir-Improta et al., 2017), but also to what extent the specific design of the physical environment supports a sustainable behavior (Gärling, 2014). The idea of addressing antecedents in evaluations of behavioral change interventions is based on the fact that changes in the motivational structure of behavior may be obtained without being observable in overt behavior.

Policies could rely on structural changes and/or psychological principles. The different psychological strategies draw on different psychological processes for behavioral change. The strategies are therefore likely to be more or less suitable and/or efficient depending on the individual’s personal and social situation, the stage of the behavior, as well as the contextual boundaries. Structural strategies draw on external incentives that make behavior with negative environmental impact more costly and behavior with positive environmental impact less costly by, for example, subsidies. Psychological strategies, however, aim to enhance motivation to engage in sustainable behavior by targeting the individual’s intrinsic motivation, for example, information and education (Steg et al., 2015). Dwyer et al. (1993) and Abrahamse
et al. (2005) present an overview of strategies in relation to the fundamental psychological principles for behavioral change differentiating between antecedent interventions aimed at influencing underlying behavioral determinants, which in turn are believed to influence behavior, and consequence strategies based on the assumption that the presence of positive or negative consequences will influence behavior.

It has been proposed that a mixture of strategies relying on antecedents and consequences would be most efficient (compare policy mixes above). This may be explained by the fact that the likelihood that the expected behavioral change resulting from a policy will occur may not only depend on the intervention as such, but also on whether the intervention matches the current stage of the individual’s behavior, for example, whether it has to be directed by external instructions or it is internalized and self-directed (Geller, 2002). The proportion of individuals that will change in response to a certain intervention has been called the plasticity of the behavior (Dietz et al., 2013). The plasticity depends on how supportive contextual factors are and on how well the intervention suits the specific individuals (Table 2).

Discussion

Designing evaluations that would better enhance transitions

If evaluation is to be a positive force in the transition toward sustainability, evaluations would have to be conducted differently. Traditionally, evaluation has been best at providing knowledge for incremental development. Stepwise continuous improvement can deliver huge cumulative effects. However, when it comes to the sustainability crises, time is short and more fundamental change is needed (EEA, 2019; IPCC, 2018).

Transitions require learning and evaluation has developed into a practice that could provide knowledge for this learning. The key aspects of the theoretical base of evaluation: methods, valuing, and enhancing use are all crucial if evaluation is going to enhance transitions. Based on the practice of evaluation, it is also clear that evaluations need to be sensitive to the context and that the focus of an evaluation largely determines its results.

If evaluations are to provide insight for long-term transitions of the sociotechnical systems for food, energy, mobility, and housing, the requirements for achieving such transitions have to be taken into account in the focusing and framing of the evaluations (B in the framework in Table 1). This does not imply that all evaluations should be focused merely on transitions, but that their role should clearly increase. Many studies have confirmed that the perception that evaluations largely focus on incremental impacts and effectiveness can clearly be established empirically (Huitema et al., 2011; Sandin et al., 2019; Schoenefeld, 2018). For an evaluation to have a focus on transitions, it implies that the evaluation should evaluate key mechanisms for transitions, that is, experimentation, visioning, and learning for acceleration; they should also be based on a system and multi-actor perspective.

A key aspect of a system perspective in an evaluation is to characterize the key elements of the present system to be able to evaluate which parts of the system are addressed by policies and programs or are supported by research and how path dependencies may prevent change. For example, a car-based transportation system consists of variety of elements, such as technologies, factories, rules and regulations, user practices, infrastructures, as well as supply and maintenance networks (Geels, 2005). In Finland, bio-fuel policies for decades targeted only the production of biofuels and had very limited impacts, but when the European Commission
(EU) introduced the Biofuels Directive (2003/30/EC) focusing on the use of biofuels the combined policy mix rapidly produced impacts (Temmes et al., 2014).

With respect to methods and data (C in Table 1), it is essential that transition theory is used to form such questions that can be empirically assessed and at the same time be relevant for transitions. The recommended approach would be to formulate program theories based on transition theory and focus on aspects that can be empirically assessed. A framework for such analyses has been presented by Linnér et al. (2012) and used to study biofuel transitions in Brazil (Maroun and Schaeffer, 2012). When goal achievement is assessed and longer time targets are taken into account, for example, climate goals for 2050, gap analyses are quite

| Theory                        | Key aspects to take into account                                                                 | Link to evaluation theory and practice                                                                 |
|-------------------------------|--------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------|
| Transition theory             | Visions and other landscape factors, experiments, (variation and selection); niches; learning; systems perspective; scales; path dependence and lock-in; and multi-actor approaches | These could be taken into account in both the focus of an evaluation and in the use of theories in the methods of the evaluations. |
| Policy analyses               | Policy instruments; side effects; framing and agenda setting; policy mixes; multi-level and polycentric governance; politics and power, legitimacy | Determining which policy instruments are in focus of the evaluation and which are part of the context. Linking the policy framing and the framing of the evaluation. Recognizing and making interests transparent in the evaluation. Linking the evaluation focus to other levels of governance from the United Nations to the local. Are there specific legitimacy issues in the policy context which should be taken into account in the evaluation? |
| The sociology of science      | Extent and variation of the research; scientific excellence; dissemination of findings; interaction with the society enabling co-evolution; alignment between research financing and societal demands; grand challenges and missions stimulating interdisciplinary research | Determining the evaluand’s context and boundaries by placing it in relation to scientific fields and societal objectives. Recognizing the role of the evaluand in knowledge production and societal missions, by unfolding agency and interconnections between different governance areas affecting the research. Recognize different outcomes in terms of codified knowledge, impact, mobility/dissemination, processes, etc. |
| Environmental psychology     | Behavior, behavioral intention, habit, perceived quality of life. Behavioral antecedents Psychological process of behavioral change Physical and social environmental factors and individual characteristics | Determining the focus of the evaluation such that it recognizes behavioral change, antecedents, and spillover effects. Recognizing relevant physical and social factors when the context of the evaluand is specified. Utilizing methods (including theory) and data that can capture behavioral change, antecedents, and motivations. Assessing the persistence of the behavioral change |
common. These analyses might be helpful to establish urgency to take additional measures, but by not being based on theory and being linear they risk giving the wrong signals. An assessment of experiments with alternative fuels or charging stations linked to a model for a transition from the fossil fuel-based car mobility system would bring more useful insights than an assessment of an extrapolation of the emission reductions achieved to the target, if achieved reductions are largely due to fuel-efficiency improvements of combustion cars.

Most evaluations focus on impacts and effectiveness, as goal achievement. Since transitions are comprehensive and affect all, the knowledge base would benefit from evaluations built on a broader set of values (D in Table 1). From a transition perspective, relevance also becomes a particularly important criterion, that is, has the evaluated policy, program, or research made any contribution to any transition toward sustainability? In addition, it is becoming more obvious that for policies to be legitimate, the transitions have to be seen as fair (EEA, 2019). Criteria related to distributions of impacts, such as equity, and about processes, for example, acceptability and transparency, are therefore also crucial.

If learning for transitions is to be enhanced the evaluations will have to be used (E in Table 1). Frequently, transition processes depend on new, or at least previously unestablished, actors, such as start-ups, new nongovernmental organizations (NGOs), or new cooperatives. It is therefore important to also facilitate use by unintended users. This requires openness and transparency as well as easy access to both evaluation results and the evaluation process.

**Utilizing social sciences in evaluations for transitions**

Utilizing transition theory in evaluations is, however, not enough if real insights for sustainability transitions are to be provided—other social sciences will have to be drawn on as well.

Science is an important input for many innovations. Sociology of science has studied how efficient universities and the scientific system more broadly are at producing knowledge and interlinking with other actors in making that knowledge usable. Sandin and Benner (2021) found that evaluations of research programs and research institutions in Sweden form a variegated knowledge base, where the results and their importance and relevance to transformational changes remain unspecified and where a synthesis of the results and a more holistic picture cannot be readily made due to the disparate natures of the evaluations. Recently, students and other stakeholders have demanded universities to do more for sustainable development. Frequently, university leaders have replied that the most important thing universities can do is to produce more knowledge, since any transition toward sustainability will have to be based on new knowledge. This argument is, however, not sufficient, since also the present sustainability crises has been caused by innovations produced with university-generated knowledge. It is therefore not self-evident that new knowledge by universities will be beneficial for sustainability transitions. Evaluations of research would therefore have to put a strong emphasis on the relevance of the efforts and research outputs for sustainability transitions. The missions included in Horizon Europe, the 8th research framework program by the EU (2019), depart from the sustainability crises and are largely founded on the ideas of transition theory (Mazzucato, 2018). Their future evaluation will thus be a real benchmark for the evaluation enterprise. The framework presented in this article is an input to this forthcoming process.

Evaluation is too often seen as just a rather technical exercise, forgetting its political implications as well as its political context. It is evident that political interests are large when the subject is large-scale societal transitions (Meadowcroft, 2009). Incorporating insights from
policy analyses into the evaluation framework has many implications. With respect to framing, it stresses the need to assess the complete policy mix, not just the instrument or program being evaluated. It is also fundamental to focus the evaluations also on the persistence of unsustainable systems and how the support to them hinders transitions. It stresses the importance of carefully examining whether vested interests and political influence over the implementation have influenced the outcomes.

Many policies aiming to support transitions aim at altering behavior or are based on the assumption that behavior does not change when new technologies are used. Eating less meat based on climate labeling would be an example of the first and unchanged driving habits with electric vehicles an example of the second. To fully understand and assess the outcome of these kinds of policies would require a thorough understanding of the target behavior ideally obtained by triangulating theoretical and methodological approaches. As illustrated above, not only the behavior per se but also the context of the behavior and its antecedents should be assessed to understand the process of change, and that counterfactual developments are designed accordingly (Johansson et al., 2020).

### The contribution to meta-evaluations and evaluation design

Meta-evaluations have become more frequent, which is good since they provide a base for developing the practice of evaluation through reflexivity. This development is partly related to the spread of the evidence movement (Hansen and Rieper, 2009), but systematic literature reviews are also becoming more frequent, due to better databases and search functions. Our approach builds on previous meta-evaluations, but it can also contribute to the practice of conducting meta-evaluation.

The framework presented (Table 1) has primarily been developed in order to plan and undertake evaluations that would better support sustainability transitions than hitherto. It can, however, also be used to make meta-evaluations of already conducted evaluations. From the meta-evaluation perspective, it resembles other frameworks (e.g. Huijtema et al., 2011; Schoenefeld, 2018), but it is more directly derived from evaluation theory. In addition, it has been developed to incorporate all essential features of evaluation practice as defined by Evert Vedung (2009). This framework can be seen as a base not only for making empirical meta-evaluations but also as a foundation for a theoretically founded discussion of justified framings of meta-evaluations.

Empirical studies applying the presented framework on new Swedish data (Neij et al., 2021; Sandin et al., 2019) have reconfirmed several previous results. Among these results are the predominance of effectiveness and impact as criteria, the quite limited use of theory, the assessment of the counterfactual development mainly limited to the perceptions of the interviewees, and the rare use of triangulation. By explicitly incorporating a transitions perspective in the meta-evaluation, Sandin et al. (2019) showed that a wider system perspective and a more thorough multi-actor approach as well as multi-actor involvement in the evaluations are needed.

### Undertaking interdisciplinary evaluations

The framework argued for in this article is putting a lot of demands on evaluations and evaluators, with respect to resources, knowledge, and skills. For this to be doable in practice, both
those commissioning and those undertaking evaluations would need to focus not just on the individual evaluation at hand but more on the stream of evaluations and a strategic approach to them (Neij et al., 2021).

Individual evaluations can support a holistic perspective and a structured learning. By strengthening coordination among evaluations, overall learning and knowledge transfer can be enhanced. This will, however, require a strategy for commissioning evaluations that would better support a holistic perspective and a structured learning. Such a strategy needs to ensure the interdisciplinary knowledge base and the skills required to conduct evaluations that would inform transitions toward sustainability.

Conclusion

There is an urgent need for rapid transitions of today’s unsustainable consumption and production systems. Such transitions would need to be directed and supported by policies and research. But in order to succeed, both policies and research efforts would have to develop based on gained experiences. There is thus a serious need for evaluations in support of transitions toward sustainability.

Traditionally, evaluation has been strongest at providing knowledge for incremental improvements. In this article, a framework (summarized in Table 1) has been developed based on evaluation theory and practice, but integrating aspects from transition theory, sociology of science, policy analyses, and environmental psychology. This framework was developed primarily in order to facilitate evaluations for transitions, but such a framework can and has also been used to assess already undertaken Swedish evaluations of policies for energy efficiency in buildings (Sandin et al., 2019). We acknowledge that the framework can be further extended and has only included some selected theories and approaches of relevance. For example, in cases where the direct interactions with ecosystems are essential (e.g. food systems), the incorporation of natural sciences such as ecology or soil chemistry might be essential.

If evaluations are to become useful in providing knowledge that informs transitions toward sustainability, evaluation practices would have to develop. The framing of evaluations would need to be more on long-term transitions than on short-term incremental change. There has to be a broad systems perspective recognizing multiple actors and comprehensive policy mixes. Nonlinear developments require theory-based approaches, where insights from transition theory, but also other social sciences are utilized. The theoretical insights, preferably incorporated through program theories, would guide both the focusing of the evaluation and the empirical material and methods used. If evaluations are to inform transitions, they would have to be used. Since transitions often involve new actors, such as start-ups or emerging NGOs, it is important that evaluation processes and outcomes are transparent, open, and easily accessible beyond intended users.

The complexity, as well as resource and knowledge implications of using the (entire) framework in individual evaluations are fully recognized. Therefore, both commissioners of evaluations and evaluators are urged to develop strategies for how entire sets of evaluations, rather than individual evaluations, could better inform transitions toward sustainability. The future well-being of humanity depends on rapid and profound transitions of consumption and production systems so that material use as well as emissions and environmental impacts are reduced. This requires a parallel transition in evaluation.
Acknowledgements

The authors would like to thank the two anonymous referees and the editor of Evaluation for their useful comments to the earlier drafts of the article. This research was developed within the project “Transition Governance”, we would like to thank the steering group: Peter Stern, Lisa Eriksson, Hanne Foss Hansen and Tobias Persson for their valuable comments during the process. An early version of the paper was presented at the 5th European Environmental Evaluators Network Forum in Copenhagen 15-16.9.2016 and at the 12th European Evaluation Society Biennial Conference in Maastricht 26-30.9.2016 we are grateful for the discussions at these events.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research that this article is based on has been funded by the Swedish Energy Agency through the project “Transition Governance” (grant 39938-1).

ORCID iDs

Per Mickwitz https://orcid.org/0000-0003-0631-1224
Sofie Sandin https://orcid.org/0000-0002-1372-6893

References

Abrahamse W, Steg L, Vlek C, et al. (2005) A review of intervention studies aimed at household energy conservation. Journal of Environmental Psychology 25: 273–91.

Alkin M (2013) Evaluation Roots: A Wider Perspective of Theorists’ Views and Influences, 2nd edn. London: SAGE.

Argyris C (1999) On Organizational Learning, 2nd edn. Malden, MA: Blackwell Publishing.

Bache I and Flinders M (2004) Multi-Level Governance and Environmental Policy. Oxford: Oxford University Press.

Bergek A, Jacobsson S, Carlsson B, et al. (2008) Analyzing the functional dynamics of technological innovation systems: A scheme of analysis. Research Policy 37(3): 407–29.

Berkhout F, Smith A and Stirling A (2004) Socio-technological regimes and transition contexts. In: Elzen B, Geels FW and Green K (eds) System Innovation and the Transition to Sustainability: Theory, Evidence and Policy. Cheltenham: Edward Elgar, 48–75.

Bijker W, Hughes T and Pinch T (eds) (1987.) The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology. Cambridge, MA: The MIT Press.

Bush V (1945) As we may think. The Atlantic, July, 101–8.

Callon M (1987) Society in the making: The study of technology as a tool for sociological analysis. In: Hughes T and Pinch T (eds) The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology. London: The MIT Press, 83–103.

Carlsson B (ed.) (1995) Technological Systems and Economic Performance: The Case of Factory Automation. Boston, MA: Kluwer Academic Publishers.

Chatterton T and Wilson C (2013) Characterising domestic energy related behaviours. In: ECREEE summer study proceedings, 15 June, Belambra Les Criques, Toulon/Hyères, France, 289–300. Available at: https://www.ecee.org/library/conference_proceedings/ecreee_Summer_Studies/2013/1-foundations-of-future-energy-policy/characterising-domestic-energy-related-behaviours/

Chen HT (2005) Practical Program Evaluation: Assessing and Improving Planning, Implementation, and Effectiveness. Thousand Oaks, CA: SAGE.

Dietz T, Stern PC and Weber EU (2013) Reducing carbon-based energy consumption through changes in household behavior. Daedalus 142: 78–89.
Donaldson A and Lipsey M (2006) Roles for theory in contemporary evaluation practice: Developing practical knowledge. In: Shaw I, Greene J and Mark M (eds) The SAGE Handbook of Evaluation. London: SAGE, 56–75.

Dosi G (1982) Technological paradigms and technological trajectories: A suggested interpretation of the determinants and directions of technical change. Research Policy 11(3): 147–62.

Dwyer WO, Leeming FC, Cobern MK, et al. (1993) Critical review of behavioural interventions to preserve the environment: Research since 1980. Environment and Behavior 25: 275–321.

European Commission (2019) Horizon Europe: The next research and innovation framework programme. Available at: https://ec.europa.eu/info/horizon-europe-next-research-and-innovation-framework-programme_en

European Environment Agency (EEA) (2019) The European Environment—State and Outlook 2020: Knowledge for Transition to a Sustainable Europe. Copenhagen: EEA.

Frickel S and Moor K (eds) (2006) Prospects and Challenges for a New Political Sociology of Science. Madison, WI: University of Wisconsin Press.

Gardner GT and Stern PC (1996) Environmental Problems and Human Behaviour. Boston, MA: Allyn & Bacon.

Gärling T (2014) Past and present environmental psychology. European Psychologist 19: 127–31.

Geels FW (2002) Technological transitions as evolutionary reconfiguration processes: A multi-level perspective and a case study. Research Policy 31: 1257–74.

Geels FW (2004) From sectoral systems of innovation to socio-technical systems Insights about dynamics and change from sociology and institutional theory. Research Policy 33: 897–920.

Geels FW (2005) The dynamics of transitions in socio-technical systems: A multi-level analysis of the transition pathway from horse-drawn carriages to automobiles (1860–1930). Technology Analysis & Strategic Management 17(4): 445–76.

Geels FW and Raven R (2006) Non-linearity and expectations in niche development trajectories: Ups and downs in Dutch biogas development (1973–2003). Technology Analysis & Strategic Management 18(3–4): 375–92.

Geller ES (2002) The challenge of increasing pro-environment behaviour. In: Bechtel R and Churchman A (eds) Handbook of Environmental Psychology. New York: Wiley, 525–40.

Gibbons M, Limoges C, Nowotny H, et al. (1994) The New Production of Knowledge: The Dynamics of Science and Research in Contemporary Societies. London: SAGE.

Gifford R (2014) Environmental Psychology, 5th edn. Colville, WA: Optimal Books.

Gifford R and Nilsson A (2014) Personal and social factors that influence pro-environmental concern and behaviour: a review. International Journal of Psychology 49: 141–57.

Gillingham K, Rapson D and Wagner G (2015) The rebound effect and energy efficiency policy. Review of Environmental Economics and Policy 10(1): 68–88.

Haas P (1992) Introduction: Epistemic communities and international policy coordination. International Organization 46(1): 1–35.

Hacker J and Pierson P (2010) Winner-Take-All Politics: How Washington Made the Rich Richer—And Turned Its Back on the Middle Class. New York: Simon & Schuster.

Hagstrom W (1965) The Scientific Community. New York: Basic Books.

Hajer M and Laws D (2006) Ordering through discourse. In: Moran M, Rein M and Goodin R (eds) The Oxford Handbook of Public Policy. Oxford: Oxford University Press, 251–68.

Hansen HF and Rieper O (2009) The evidence movement: The development and consequences of methodologies in review practices. Evaluation 15(2): 141–63.

HEFCE (2016) Publication patterns in research underpinning impact in REF2014: A report to HEFCE by Digital Science. Available at: https://dera.ioe.ac.uk/26933/1/2016_refimpact.pdf

Hogg D, Elliott T, Elliott L, et al. (2016) Study on Assessing the Environmental Fiscal Reform Potential for the EU28. Luxembourg: Publications Office of the European Union. Available at:
Hoogerwerf A (1990) Reconstructing policy theory. *Evaluation and Program Planning* 13(3): 285–91.

Hughes TP (1983) *Networks of Power: Electrification in Western Society 1880-1930*. Baltimore, MD: The Johns Hopkins University Press.

Huijts NMA, Molina EJE and Steg L (2012) Psychological factors influencing sustainable energy technology acceptance: A review-based comprehensive framework. *Renewable and Sustainable Energy Reviews* 16: 525–31.

Huitema D, Jordan A, Massey E, et al. (2011) The evaluation of climate policy: Theory and emerging practice in Europe. *Policy Sciences* 44(2): 179–98.

IPCC (2018) Global warming of 1.5°C. An IPCC special report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty. World Meteorological Organization, Geneva.

Johansson M, Gentile N and Neij L (2021) Evaluating energy efficient behaviour in the built environment - in what way can environmental psychology strengthen the current evaluation practice in the Nordic countries? *Energy Efficiency*, forthcoming. DOI: 10.1007/s12053-021-09938-z

Jørgensen U (2012) Mapping and navigating transitions—The multi-level perspective compared with arenas of development. *Research Policy* 41: 996–1010.

Kaiser T (2017) *The European Research Council*. Cambridge: Polity.

Kemp R (1994) Technology and the transition to environmental sustainability. *Futures* 26: 1023–46.

Kern F, Kivimaa P and Martiskainen M (2017) Policy packaging or policy patching? The development of complex energy efficiency policy mixes. *Energy Research & Social Science* 23: 11–25.

Kivimaa P and Kern F (2016) Creative destruction or mere niche support? Innovation policy mixes for sustainability transitions. *Research Policy* 45(1): 205–17.

Lenoir-Improta R, Devine-Wright P, Pinheiro JQ, et al. (2017) Energy issues: Psychological aspects. In Fleury-Bahi G, Pol E and Navarro O (eds) *Handbook of Environmental Psychology and Quality of Life Research*. Cham: Springer, 543–57.

Linnér B-O, Mickwitz P and Román M (2012) Reducing greenhouse gas emissions through development policies: A framework for analysing policy interventions. *Climate and Development* 4(3): 175–86.

Loorbach DA (2007) *Transition Management: New Mode of Governance for Sustainable Development* [Transitiemanagement; nieuwe vorm van governance voor duurzame ontwikkeling]. Utrecht: International Books.

Maroun C and Schaeffer R (2012) Emulating new policy goals into past successes: Greenhouse gas emissions mitigation as a side effect of biofuels programmes in Brazil. *Climate and Development* 4(3): 187–98.

Martin B (2016) R&D policy instruments: A critical review of what we do and don’t know. *Industry and Innovation* 23(2): 157–76.

Mazzucato M (2018) *Mission-Oriented Research & Innovation in the European Union: A Problem-Solving Approach to Fuel Innovation-Led Growth*. Brussels: European Commission.

Meadowcroft J (2009) What about the politics? Sustainable development, transition management, and long term energy transitions. *Policy Sciences* 42(4): 323–40.

Merton R (1973) *The Sociology of Science: Theoretical and Empirical Investigations*. Chicago, IL: University of Chicago Press.

Mickwitz P. 2003. A framework for evaluating environmental policy instruments: Context and key concepts. *Evaluation* 9(4): 415–36.

Moed H (2006) *Bibliometric Rankings of World Universities* (CWTS Report 2006-01). Leiden: Centre for Science and Technology Studies (CWTS), Leiden University.
Moezzi M and Janda KB (2014) From “if only” to “social potential” in schemes to reduce building energy use. *Energy Research and Social Science* 1: 30–40.

Moser G (2009) Quality of life and sustainability: Toward person-environment congruity. *Journal of Environmental Psychology* 29: 351–57.

Neij L, Sandin S, Benner M, et al. (2021) Bolstering a transition for a more sustainable energy system: A transformative approach to evaluations of energy efficiency in buildings. *Energy Research & Social Science* 72: 101864.

Patton MQ (2008) *Utilization-Focused Evaluation*, 4th edn. Thousand Oaks, CA: SAGE.

Pestre D (2003) Regimes of knowledge production in society: Towards a more political and social reading. *Minerva* 41(3): 245–61.

Peters M, Schneider M, Gresshaber T, et al. (2012) The impact of technology-push and demand-pull policies on technical change: Does the locus of policies matter? *Research Policy* 41(8): 1296–308.

Popper K (2003) *The Open Society and Its Enemies, Vol. Two: Hegel and Marx*. London: Routledge Classics.

Pressman J and Wildavsky A (1984) *Implementation: How Great Expectations in Washington Are Dashed in Oakland*, Third expanded edn. Berkeley, CA: University of California Press.

Raven R (2007) Niche accumulation and hybridisation strategies in transition processes towards a sustainable energy system: An assessment of differences and pitfalls. *Energy Policy* 35(4): 2390–400.

Raven R, Schot JW and Berkhout F (2012) Space and scale in socio-technical transitions. *Environmental Innovation and Societal Transitions* 4: 63–78.

Rip A and Kemp R (1998) Technological change. In: Rayner S and Malone EL (eds) *Human Choice and Climate Change* (Vol. 2). Columbus, OH: Battelle Press, 327–99.

Rogge K and Reichardt K (2016) Policy mixes for sustainability transitions: An extended concept and framework for analysis. *Research Policy* 45(8): 1620–35.

Rotmans J, Kemp R and Van Asselt M (2001) More evolution than revolution: Transition management in public policy. *Foresight* 3(1): 15–31.

Sabatier P and Jenkins-Smith H (1993) *Policy Change and Learning: An Advocacy Coalition Approach*. Boulder, CO; San Francisco, CA; Oxford: Westview Press.

Sandin S and Benner M (2021) Research evaluations for an energy transition? Insights from a review of Swedish research evaluation reports, under review.

Sandin S, Neij L and Mickwitz P (2019) Transition governance for energy efficiency: Insights from a systematic review of Swedish policy evaluation practices. *Energy, Sustainability and Society* 9(17): 1–18.

Schmookler J (1962) Changes in industry and in the state of knowledge as determinants of industrial invention. In: Universities-National Bureau Committee for Economic Research, Committee on Economic Growth of the Social Science Research Council (ed) *The Rate and Direction of Inventive Activity: Economic and Social Factors*. Cambridge, MA: National Bureau of Economic Research, Inc, 195–232.

Schoenefeld J (2018) *Evaluation in polycentric governance systems: Climate change policy in the European Union*. Doctoral Thesis, University of East Anglia. Available at: https://ueaeprints.uea.ac.uk/69535/

Schoenefeld J and Jordan A (2017) Governing policy evaluation? Towards a new typology. *Evaluation* 23(3): 274–93.

Schot JW and Geels FW (2008) Strategic Niche Management and sustainable innovation journeys: Theory, findings, research agenda and policy. *Technology Analysis & Strategic Management* 20(5): 537–54.

Schot JW, Hoogma R and Elzen B (1994) Strategies for shifting technological systems. *Futures* 26: 1060–76.
Schuitema G and Bergstad Jakobsson C (2013) Acceptability of environmental policies. In: Steg L, van den Berg AE and de Groot JIM (eds) Environmental Psychology an Introduction. The British Psychological Society. Toronto, Canada: John Wiley & Sons, 295–306.

Scriven M (1991) Evaluation Thesaurus, 4th edn. Newbury Park, CA: SAGE.

Shadish WR, Cook TD and Leviton LC (1995) Foundations of Program Evaluation: Theories of Practice. Newbury Park, CA: SAGE.

Skogstad GD (ed.) (2011) Policy Paradigms, Transnationalism, and Domestic Politics. Toronto, ON, Canada: University of Toronto Press.

Smith A, Stirling A and Berkhout F (2005) The governance of sustainable socio-technical transitions. Research Policy 34(10): 1491–510.

Steg L, Perlaviciute G and van der Wer E (2015) Understanding the human dimensions of a sustainable energy transition. Frontiers in Psychology 6: 805.

Stern P, Enberg J, Jondell Assbring M, et al. (2013) Metautvärdering av utvärderingar genomförda på Energimyndighetens uppdrag 2000-2012 [Meta-evaluation of evaluations for the Swedish Energy Agency 2000-2012]. Stockholm: Faugert & Co Utvärdering AB.

Streicher M (2006) An interview with Helga Nowotny: The European Research Council (ERC): Fostering basic research in Europe. Bridges 10: 29.

Temmes A, Virkamäki V, Kivimaa P, et al. (2014) Innovation Policy Options for Sustainability Transitions in Finnish Transport (Tekes Reviews 306/2014). Helsinki: Teknologian kehittämisskeskus.

Truelove HB, Carrico AR, Weber EU, et al. (2014) Positive and negative spillover of pro-environmental behavior: An integrative review and theoretical framework. Global Environmental Change 29: 127–38.

Turnheim B, Berkhout F, Geels F, et al. (2015) Evaluating sustainability transitions pathways: Bridging analytical approaches to address governance challenges. Global Environmental Change 35: 239–53.

Turnheim B, Kivimaa P and Berkhout F (2018) Innovating Climate Governance: Moving Beyond Experiments. Cambridge: Cambridge University Press.

Valovirta V (2002) Evaluation utilization as argumentation. Evaluation 8(1): 60–80.

van den Bergh J, Truffer B and Kallis G (2011) Environmental innovation and societal transitions: Introduction and overview. Environmental Innovation and Societal Transitions 1(1): 1–23.

Vedung E (1997) Public Policy and Program Evaluation. New Brunswick: Transaction Publishers.

Vedung E (1998) Policy instruments: Typologies and theories. In: Bemelmans-Videc ML, Rist R and Vedung V (eds) Carrots, Sticks, & Sermons: Policy Instruments & Their Evaluation. New Brunswick: Transaction Publishers, 21–58.

Vedung E (2009) Utvärdering I politik och förvaltning, 3:e upplagan. Lund: Studentliteratur.

Vedung E (2010) Four waves of evaluation diffusion. Evaluation 16(3): 263–77.

Zuckerman H (1977) Scientific Elite: Nobel Laureates in the United States. New York: Free Press.

Per Mickwitz is a Professor of environmental policy at the International Institute for Industrial Environmental Economics and since 2021 pro vice-chancellor for research, sustainability and campus development at Lund university. His research has focused on environmental and sustainability policy, how politics can be evaluated and how innovations occur. He has a particular interest in how system transitions towards sustainability can be supported through policies, research and evaluation.

Lena Neij, Professor at the International Institute for Industrial Environmental Economics, Lund university, focuses in her research on the dynamics of the energy system and the analysis of governance and policies for a more sustainable development. The research, which has been system based and interdisciplinary, has covered energy for sustainable development in general, with a focus on renewable energy technologies and energy efficiency in buildings.
Maria Johansson, Professor in environmental psychology researching human-environment transactions from the individual’s perspective in urban and natural settings. One research interest concerns environmental design and energy efficiency behaviour in the built environment.

Mats Benner is a Professor and Dean of Lund University School of Economics and Management. He studies how governments set priorities for research and innovation, and how they govern universities.

Sofie Sandin is a PhD candidate at the International Institute for Industrial Environmental Economics at Lund University. Her research focuses on transformative evaluation approaches for research and policy incentives aimed at energy efficiency in buildings.