



Philosophy of sustainability experimentation _ experimental legacy, normativity and transfer of evidence

Milutin Stojanovic¹

Received: 17 June 2020 / Accepted: 20 May 2021 / Published online: 31 July 2021
© The Author(s) 2021, corrected publication 2021

Abstract

The recent proliferation of types and accounts of experimentation in sustainability science still lacks philosophical reflection. The present paper introduces this burgeoning topic to the philosophy of science by identifying key notions and dynamics in sustainability experimentation, by discussing taxonomies of sustainability experimentation and by focusing on barriers to the transfer of evidence. It integrates three topics: the philosophy of experimentation; the sustainability science literature on experimentation; and discussions on values in science coming from the general philosophy of science, the social sciences, and sustainability science. The aim is to improve understanding of how sustainability experimentation has evolved, from a broader picture of the history and philosophy of science, with a specific focus on understanding evidence production and how evidence traveling in and from sustainability experiments can be improved, particularly in the context of complex and pervasive normative commitments of the research. By engaging in these topics, this research is one of the first philosophical accounts of sustainability experimentation, contributing both to the knowledge on specific philosophies of science and to the further development of an evidence-based sustainability science through a better understanding of the barriers to more relevant and usable knowledge.

This article belongs to the Topical Collection: EPSA2019: Selected papers from the biennial conference in Geneva

Guest Editors: Anouk Barberousse, Richard Dawid, Marcel Weber

✉ Milutin Stojanovic
milutin.stojanovic@helsinki.fi

¹ Helsinki Institute of Sustainability Science (HELSUS) and Practical Philosophy, University of Helsinki, Helsinki, Finland

1 Introduction: grounds for the philosophy of sustainability experimentation

Evidence is the hallmark of science, and experimentation¹ is the main tool for producing evidence in a range of scientific fields (Hacking, 1983; Radder, 2003). Experimentation is the main way to obtain clues for causal inference, test the veracity of scientific theories, eliminate alternative explanations, and is also an important driver of innovation and novel solutions to practical problems (Feynman et al., 1963; Thye, 2014). The existing literature on experiments and experimentation has not yet produced a unified definition of what experiments are, except that they are initiatives that deviate from currently normalized practices (Hilden et al., 2017). Instead of a unified definition, different scientific disciplines have highlighted various aspects of experimentation, thereby contributing to John Dewey's idea of a pluralistic 'experimental society' or 'culture of experimentation' (VanderVeen, 2011).² This idea grew during the twentieth century in the context of the social sciences, producing concepts such as – 'society as laboratory', 'risk society', 'collective experimentation', etc.

Experiments are crucial for sustainability science because they allow researchers to produce evidence about both the causes of sustainability problems, and about the effectiveness of their solutions (Caniglia et al., 2017), given the widely acknowledged overarching goal of sustainability science – the transformation of society towards new practices and organizational structures (Kates et al., 2001; Rotmans et al., 2001). Sustainability experimentation is also a central driver of change in human-natural systems, the main aim of which is to provide knowledge about new solutions to current sustainability problems (Luederitz et al., 2016; Ansell & Barthenberger, 2016; Jalas et al., 2017). In testing and inciting fundamental processes of change of practices, values and social order, we are witnessing a proliferation of known forms and types of experimentation (e.g., living labs, transformation labs, real-world laboratories, etc.).

The diversity of experiments in the burgeoning field of sustainability science (SS) is creating an increasingly complex research area and the literature reporting on sustainability experimentation is rapidly growing (e.g., Sengers et al., 2019; Kivimaa et al., 2017a, b; Laakso et al., 2017; Caniglia et al., 2017; Weiland et al., 2017; Heiligenberg, 2017; Schapke, 2018). Although some recent systemic studies of experimentation in SS make the connection to the philosophy of science (e.g., to classical Baconian experimentation) as a point of reference for developing a somewhat broader view of experiments (e.g., Caniglia et al., 2017; Weiland et al., 2017), the recent proliferation of perspectives on sustainability experimentation needs a deeper, comprehensive philosophical account of the key notions and major dynamics. In particular, we need to develop a

¹ Throughout this article, we use 'experiments' to refer to the process of performing a scientific procedure, and 'experimentation' to refer to scientific practice, including its scientific approach, background knowledge, broader context and outputs.

² This also influences the contemporary science-policy discourse – the Finnish Government Program of 2015, for example, explicitly aspires to create a 'culture of experimentation' (Government Programme, 2015).

better understanding of the philosophical motives, researchers' roles, responsibilities and possible outcomes of experimentation, including particularly experimental assumptions, learning, and how evidence travels in SS (transferability of experimental results).

In seeking to bring greater coherence to the literature on sustainability experimentation and while trying to identify barriers and limitations and ensure its effectiveness and avoid pitfalls, I investigate the key philosophical notions and dynamics of sustainability experimentation, highlighting comparisons with other experimental fields, and addressing the theoretical underpinnings of sustainability experimentation including the critical question regarding the roles of values in SS. For this purpose, it seems fruitful to view sustainability experimentation in the context of the history and philosophy of experimentation and to draw conclusions from this rich and significantly developed field in the philosophy of science. This study draws upon comparisons with philosophy of experimentation in sciences with a historically contingent object of study, and lessons from these sciences regarding the theory-evidence relationship, (philosophical) motives for field experimentation, and handling of normativity.

Specifically I track the evolution of experimentation in ecology, evolutionary biology and the various social sciences, and shed light on sustainability experimentation from this perspective. The specific focus is on how the liberalization of experimental control is enacted through the aim of increasing the external validity of experiments and the problems this raises. Investigating taxonomies and systematizations coming from SS and discussing their theoretical motives, I connect the observed low generalizability and transferability of results of sustainability experimentation with undisclosed roles of values, both in driving the participants' behavior, and in framing the research problem and results of the experiments. We connect this problem with discussion on the role of values in science (revolving the argument of inductive risk in hypothesis testing) (Douglas, 2000) and highlight the specific form this discussion takes in SS. By identifying implicit normative assumptions driving the experiments and drawing attention to the problem of values, I argue for increased transparency of normative assumptions and orientations underlying the research as a means to delineate the evidence-sharing classes in SS and thereby making the transferability problem manageable. From there the paper turns to consider the role of societal learning in and from experiments, including the sharing and use of knowledge, suggesting how research approaches might increase the potential for transferring evidence. I conclude with a discussion of new ways to move forward in the research on experiments, arguing that in depth study and classification of normativity can facilitate the advantages coming from the closer model-world relation in real-world sustainability experiments and increase the external validity and transferability of the results.

Engaging in these topics, this research is one of the first philosophical accounts of sustainability experimentation and it will contribute to the further development of evidence-based SS, in the framework of philosophical contributions to SS (Nagatsu et al., 2020), through a better understanding how evidence is produced in SS, how it travels, and what the barriers for a more relevant and usable knowledge are. From the perspective of philosophy, this paper enriches the philosophy of science literature by adding to the philosophical knowledge on the specific sciences, namely SS.

This would contribute to Arthur Fine's project of specification of philosophy of science that he outlined for the field some thirty years ago (Kitcher, 2019).

2 Methodology

Our starting position is a theoretically messy and underdeveloped situation in the philosophy of sustainability experimentation, relative to accounts of experimentation in other scientific fields. This study connects three literatures: the philosophy of experimentation in life and social sciences as a part of general philosophy of science; the review and meta-review literature from SS on experimentation for sustainability solutions (e.g. Luederitz et al., 2016; Sengers et al., 2019; Caniglia et al., 2017; Weiland et al., 2017; Kivimaa et al., 2017a, b; Heiligenberg et al., 2017); and the literature on normativity and the role of values in science from the philosophy of science, coupled with the literature on context dependency and ethical values in the social and sustainability sciences.

Section (2) provides an overview of how experimentation has evolved in biology and the social sciences, track the gradual liberalization of experimental control leading to SS, and discuss the role of this liberalization in increasing the external validity of experiments in these fields. Section (3) introduces sustainability experimentation, discuss its taxonomies, their motives and limitations, and theoretical grounds for pluralism of experimental modes characteristic for the field. Section (4), based on the philosophy of social sciences and the literature on values in science, discusses a key barrier for transferability and generalizability of results in SS (namely, normativity inherent in SS) and explore options for ordering the field (through explicating normative orientations of experiments) and increasing evidence traveling. The concluding Sect. (5) summarizes the main argument.

3 Evolution of experimentation _ deviations from the classical account

Since Ian Hacking's seminal *Representing and Intervening* (1983) much of the philosophical work that has been done on experimentation has restricted itself to the physical sciences and to a lesser extent cell biology, focusing on theory-testing in controlled experimental settings. A main interest of this section is in thinking about the slow deviation of experimentation in the life and social sciences from this classical account,³ particularly regarding control of experimental conditions, and subsequently, the extent to which this evolution of experimentation sets the stage for understanding and philosophically unraveling sustainability experimentation. Nevertheless, I hope to elucidate some significant underlying differences between the physical, life, and social sciences in terms of the role of experiments in these fields.

³ In the description of the classical account my aim has not been to give an all-encompassing account, but for the sake of brevity, to focus on characteristics philosophically relevant to the discussion of SS. For more details on the classical account see (Radder, 2003).

In summary, this section will argue how experimentation in these different fields has evolved organically towards experimentation in SS and how their philosophies contribute to understanding it. However, as we will see in the next section, philosophies of experimentation in these fields fall short of capturing unique distinguishing characteristics of sustainability experimentation.

3.1 Field experiments in ecology and evolutionary biology

In some natural but at the same time historical sciences (see Cleland, 2001), namely ecology and evolutionary biology, characteristics of experiments have already started to deviate from the classical account (Wilson, 2009). The main philosophical differences stem from the fact that many of the experiments in ecology and evolutionary biology are done in the field, rather than in the laboratory. In the former case, the natural environment constitutes a relatively extensive part of the experimental setting and is not under the experimenter's direct control (e.g., the acidity of water in a natural lake). Field experimentation introduces the gradient of manipulation of experimental conditions and experimental ecologists accordingly distinguish between laboratory, field, and natural experiments (i.e., naturally occurring perturbations in the target system – with no control over the conditions) (Diamond, 1986). Starting the 1950s and culminating in the 1960s, community ecology field experiments had a profound influence on many other scientific fields and gave support to a methodological trend in favor of field experimentation (Grodwohl et al., 2018).

Field experiments have both advantages and disadvantages when compared to laboratory and natural experiments. The change in the experimental setting, no longer being artificially isolated in the laboratory but now embedded into the natural environment, is primarily made with the motive of strengthening the connection between the experimenter's model and the targeted natural system, thereby enhancing the external validity (i.e. generalizability) of experimental findings (Gerber & Green, 2011). Here, the distance between the experimental system and the natural system is diminished (almost by definition), facilitating how experimental results pertain to actual systems in nature (Wilson, 2009). Compared to natural experiments, field experiments introduce the randomization of key variables, crucial for causal inference and enhancing external validity in fields in which units of observation (e.g., individuals, groups, institutions, states) need to be randomly assigned to treatment and control groups.

However, not all field experiments necessarily exhibit stronger external validity, since such validity depends on (uncontrollable) background factors and contexts (Dipboye & Flanagan, 1979). In other words, due to their context dependency (as real-world settings differ dramatically), it may be difficult (or even impossible) to generalize the results of field experiments to systems other than those that are being directly studied (Grodwohl et al., 2018) – a problem which will become particularly salient in sustainability experiments, due to their contextual and normative complexities (see §4). As a consequence, there is often a trade-off between internal and external validity, and both are rarely captured in a single experiment – making replication

crucial, together with some combination of laboratory, natural and field studies, to obtain stable, repeatable and generalizable results in the field (an approach later taken in social psychology, economics and other social science fields).

Studying experimentation in evolutionary biology, and following the above developments in ecology (Diamond, 1986), Robert Brandon (1996) distinguishes two dimensions along which experiments vary:

1. manipulation of experimental conditions; and
2. hypothesis testing,

creating a classification of studies in evolutionary biology that can also be applied to experiments in other disciplines, including SS. The degree of manipulation involved (1) reflects the increasing prominence of field experimentation and related relaxation of experimental control from the ideal of classical experimentation. The hypothesis testing dimension (2), embodying differences in the role and development of the background theory, oscillates from experimenting with the aim of testing a hypothesis or theory on one hand, to measuring a parameter or simply describing some important aspect of nature on the other (e.g. exploratory experimentation (Stojanovic, 2013)). These two dimensions create a six-fold classification of experiments,⁴ with many intermediate forms in between. (This classification will serve as the basis for classifying experiments in SS in §3).

The important philosophical idea behind this classification is that highly-manipulative hypothesis testing (traditionally associated with laboratory experiments in the physical sciences) should not be considered ‘the most experimental’ form which should constitute the basis of any experimental science – an ideal that was ubiquitously present in traditional physics-based accounts of science and often mixed with the idea of ‘mature science’. Instead, the reasons for the appropriateness of a particular type are markedly different across the sciences and topics, and many forms of experiments may be appropriate, depending on the different scientific purposes and contexts in which experimentation takes place. For example, although non-hypothesis-driven experimentation is present even in physics due to the importance of measuring physical properties (particularly constant physical values), the appropriateness of this type in other sciences may vary. For example, the high value of non-hypothesis testing experiments in biology is related to the object of this science being uniquely distributed in time and the peculiar role of general theory (i.e., difficulties in deriving phenomenological models from the abstract framework of the Darwinian evolutionary theory, and high dependency (relatively to physics) on contextual factors).⁵ Much of this reasoning transfers to other sciences studying historical phenomena, notably social structures and social psychological constructs (e.g.

⁴ Types of manipulation (laboratory; field; natural) across different aims (hypothesis testing; measurement/exploratory experimentation).

⁵ Although, as Cartwright (1999) demonstrates, even theoretical physics is dominated by the high autonomy of phenomenological models and semi-empirical methods in their construction – case in point are quantum atomic nucleus models.

sociology, social psychology, but also SS), contributing to greater autonomy of experimental models in these sciences and liberalization of the methodologies and types of experiment.

3.2 Social experimentation

In the latter part of the twentieth century, following the trend instigated by community ecology experiments, many social sciences started moving experiments from the laboratory to the real world (Thye, 2014). Field experimentation first became the norm in psychology (Mandler, 2007), and later among sociologists (Hausman & Wise, 1985) and economists (Levitt & List, 2008), primarily based on the concerns about generalizability of experimental results (Benz & Meier, 2006).

Experiments in the social sciences that feature this drive towards more realistic models (by reducing control and increasing inclusion of environmental or contextual factors) are often performed in ways that participants in a field experiment are unaware that the events they experience are an experiment, or they know there is an experiment but they do not know what is being tested or possess false information regarding it. Philosophically important characteristic of these forms of experimentation is extension of experimental risks into the wider society (Krohn & Weyer, 1994; also, Latour, 2011), including unexpected or undesirable outcomes which previously remained within the controllable conditions of the laboratory (Guggenheim, 2012; Weiland et al., 2017). Accordingly, as scholars from science and technology studies have argued for decades, “ideas of ‘laboratory’ and ‘experiment’ have ventured outside of their natural science confines and invaded society at large (Krohn & Weyer, 1994), thus blurring the strict lines between the privileged scientific knowledge and the pragmatic knowledge of everyday life (Karvonen & Van Heur, 2014).

As the usual aim of field researchers is to learn how to modify behavior (solution-orientedness) that has proven to be recalcitrant in the past (e.g. poor school performance, drug abuse, unemployment, or unhealthy lifestyles) as opposed to testing a theoretical proposition about unidimensional causes and effects, the social sciences introduce a characteristic normative dimension to experimental research, prescribing a desirable state of the target system and directing the intervention. Normativity of the research approaches in the social sciences puts specific limits on the transfer of evidence as various schools adhere to often incommensurable research assumptions, thereby fragmenting the discipline (Graeber, 2001) and proliferating modes of experimentation. (I will focus on the fragmentation problem in §4.)

This section draws attention to the evolution of experimentation across the sciences, commonly underrepresented in the philosophical literature, and at the same time prepared the theoretical ground for understanding sustainability experimentation. In the next section, we will see how the peculiarities of the scientific object and of the research approach play out in SS (including the state of background theory). In SS, a high degree of manipulation of experimental conditions (as in biology and the social sciences), is often undesirable and inappropriate for answering interesting and important questions from the real world. Also in SS, a general theory is characteristically missing due

to the inter-disciplinary and trans-disciplinary aspirations of the field. Instead, SS puts forwards a specific solution-oriented approach, and further liberalizes experimental control focusing on case-specific real-world experimentation, encountering a specific continuation of the above problems for evidence traveling.

4 Experimentation in sustainability science

4.1 Conceptual and theoretical underpinnings of sustainability experimentation

In sustainability science, researchers examine human-natural interactions and dynamics which are characterized by complexity and uncertainty (Clark, 2007; Mitchell, 2009; Nowotny, 2015; Abson et al., 2016) and openly embrace a system perspective on the multiple and interacting social, economic, cultural and ecological factors that lead to the emergence of sustainability problems (Kates et al., 2001; Wiek et al., 2012). Although many SS experiments remain very close to the classical approach, many are carried out in real-world settings (Schapke et al., 2018), often situated in specific social, cultural, and geographical contexts (Karvonen & van Heur, 2014). These contexts include the participation of social actors in the experimental design and governance engaged around the aim to develop alternative socio-ecological-technological visions and opportunities (Schot & Geels, 2008; Sengers et al., 2019). Real-world sustainability experiments are heavily and explicitly engaged in value detection and co-creation, involving diverse scientific and social actors in the process of knowledge production (Lang et al., 2012). The focus is on social learning and empowerment as central goals of scientific work in this field (Kincaid et al., 2007; Lang et al., 2012) (although proximate drivers such as stakeholder pressures and funding constraints may have the upper hand in many cases). In these aims, inter- and trans-disciplinary research is employed with the double objectives of: 1) exploring the complex interactions and limitations of socio-ecological systems, and, building on that, 2) informing and inciting sustainability transformations (Grin et al., 2010; Wiek et al., 2012).

Sustainability transformations is a term intended to refer to long-term, multi-dimensional and fundamental processes of change through which established socio-technical systems shift to more sustainable modes of production, consumption, and social order and systems of values (Rotmans et al., 2001; Shove & Walker, 2010). Sustainability-oriented experimentation is seen as a main instigator and driver of these transformation processes and the literature on sustainability experiments has been growing rapidly in recent years (e.g. Jalas et al., 2017; Luederitz et al., 2016; Sengers et al., 2019; Weiland et al., 2017). Sustainability experiments range from testing specific technological or organizational solutions (e.g., testing new standards in a specific sector), to producing evidence on large-scale societal transitions: like alternative economic organizational structures in eco-industrial parks, or alternative patterns of consumption and mobility in cities (e.g., Grin et al., 2010; Laakso et al., 2017). However, despite the broad emphasis

on real-world contexts and far-reaching impacts, examples are typically carried out in micro contexts (i.e., local, group or even individual contexts).

When speaking about the roles played by experiments in producing evidence in SS, it is important to note the heterogeneous status of background theory. Due to the multi-disciplinary (MD) approach, the background theory used in sustainability experimentation is relatively fragmented, consisting of a range of disciplinary elements and there are fundamental difficulties when constructing models piecemeal in the course of MD collaborations (MacLeod & Nagatsu, 2018). In light of these peculiarities regarding the complex, context-dependent and value laden object of research and background theory heterogeneous status and constricted role, traditional accounts of experimentation cannot always capture the way sustainability experiments produce evidence. Although important similarities and overlaps exist with experimentation in other branches of science, the types most distinctive in SS, notably sustainability transformation experiments, are markedly left without an explanation of experiment-evidence connection and without a proper understanding of the limits of generalizability or transferability of the evidence produced (Adler et al., 2018). Additionally, because traditional accounts of experimentation are founded on an ideal of objectivity that sees scientific knowledge as universal, value-free, and independent of contexts (Kincaid et al., 2007; Mitchell, 2009; Nowotny, 2015), sustainability experimentation needs its own unique philosophical account. It requires development of an alternative value- and practice-based framework to navigate intricate scientific pathways of the current stage of SS.

4.2 Changes in the central dimensions of experimentation and taxonomy of sustainability experiments

Weiland et al. (2017) identified the key departures in sustainability experimentation from the traditional accounts of experimentation: 1) the aim of knowledge production, 2) the roles of experimenters and participants, and 3) the unpredictability of outcomes. Both (2) and (3) are essentially connected with the complex context-dependent object of SS and correlated with experimentation in the social sciences. The roles of experimenters and participants in SS (2) expand to include participatory observation techniques (in which experimenters are part of the social setting under study), and stakeholder co-creation of experimental design and governance (where diverse social actors are active experiment designers and governors). As a result, participation of different stakeholders in sustainability research is accordingly problematized, creating a need to clarify the roles scientists and society play in sustainability experimentation (including how research conflicts are resolved) (Jalas et al., 2017; Krohn & Weyer, 1994). As sustainability transformations experiments are ecological field experiments on human societies, the unpredictability of outcomes (3) is now transferred from controlled laboratory environments to society at large, including the exposure of society to foreseen and unforeseen, and positive and negative outcomes (Guggenheim, 2012; Weiland et al., 2017). However, this destabilization of standard cognitive procedures can be viewed as options for new

knowledge production and societal learning (Luederitz et al., 2016). In other words, focus on limiting unpredictability is deliberately traded off in favor of exploring societal options (particularly in experiments aiming for production of transformation knowledge).

The aim of experimental research in SS (1) can be broken down to the following knowledge objectives: (1.a) production of evidence on the causal links and dynamics of socio-ecological systems (**system knowledge**); (1.b) understanding what a system should be like (i.e., knowledge about desirable sustainability targets) (**target knowledge**) (Brown, 1997; Guggenheim, 2012); and evidence on how to transform the system (i.e., testing opportunities for change of technological or social trajectories) (**transformation knowledge**) (ProClim, 1997; Grunwald, 2004; Weiland et al., 2017). Understanding causal links and dynamics of phenomena (production of system knowledge (1.a)) is part and parcel of classical experimentation from its inception in the natural sciences. Evidence about desirable state of the system under study (target knowledge (1.b)), as well as evidence about mechanisms for transforming the system to the desirable state (transformation knowledge (1.c)), were introduced with the advent of social experimentation (Brown, 1997), but remained largely in the background due to positivistic ambitions in disciplines such as medicine, psychology, sociology, or economics. With the advent of SS in the last few decades, value laden, normative political goals regarding improving the situation inadvertently became impossible to shun completely. However, explicit discussion of normativity⁶ in sustainability experiments, including value judgments they presuppose (in particular related to what characterizes ‘sustainability’ or ‘sustainable development’) is still mostly missing from the literature (Schneider et al., 2019).

Based on the shared value-ladenness of target and transformation knowledge, it is possible to describe the aims of experimental research (1) in different terms. Caniglia et al. (2017) identify non-epistemic normative goals of sustainability experimentation as a single dimension (conflating target-oriented and transformation-oriented experimentation), as both are related to actionable knowledge. Rephrasing the terminology somewhat, experimental aims (1) are theoretically separated on production of evidence about causal system knowledge regarding sustainability problems (i.e. complex interactions of natural and social systems) and about the effectiveness of solutions – actionable knowledge. The latter includes sustainability targets that actors should be striving for, and the process of achieving the change of social and technological trajectories) (Miller et al., 2014; Clark et al., 2016, Caniglia et al., 2017). However, the system and actionable knowledge are rarely clearly distinguished in sustainability research (Wiek et al., 2012).

Resembling the dimensions of hypothesis testing and of manipulation of experimental conditions introduced by ecologists and evolutionary biologists a few decades earlier (Brandon, 1996; Diamond, 1986) (see Sect. 2.1), Caniglia

⁶ Following Schneider et al. (2019), we use the term ‘normative’ as an umbrella term for all issues concerning values, understood as reference points for evaluating things as good or bad. Values are “conceptions of desirable”, influencing the choices people make between different possible courses of action (Graeber, 2001) and are rationally and emotionally binding sources of orientation and motivation. It may be important to note that there is no comprehensive theory of value across the sciences (Graeber, 2001).

Table 1 Taxonomy of sustainability experimentation. Integrated from Camiglia et al. (2017) and Weiland et al. (2017)

B: Aim of evidence production		A: Degree of manipulation		
		Full-control	Participatory-control (co-creation)	No-control
Causal system knowledge	Laboratory experiments	Adaptive experimentation (uncommon)		Implicit, natural, quasi-natural experiments
Normative, target knowledge	N/A	real-world value co-creation experiments, living labs		Studies in strategic niche management and reflexive governance
Actionable transformation knowledge	Testing innovations and behavior in living labs	transformation labs, real-world labs, urban sustainability labs		Studies in strategic niche management and transition management

et al. (2017) make a sixfold taxonomy of sustainability experimentation (types of control across types of knowledge aims – see Table 1 for the amended taxonomy) – analogous to Brandon’s taxonomy of experiments in biology discussed in 2.1, but according to experimental aims instead of an hypothesis testing dimension.

The degree of manipulation varies across types of knowledge and aims of experiments, decreasing characteristically in the production of normative and actionable knowledge. Although some types of experiment happen both in relatively highly-controlled and participatory settings (e.g. living labs), controllability of sustainability transformation experiments is generally held to be quite low due to the need to share experiment design and governing capacities with the relevant stakeholders and include their values and assumptions. Drawing on Brandon’s taxonomy described in Sect. 2.1, we can observe how the ‘hypothesis-testing’ dimension turned into ‘aims of evidence production’, due to the decreased role of general theory in SS and its unique focus on actionable, solution-oriented knowledge (Miller et al., 2014), including rich sustainability-specific normative contexts. Experimental aims in SS are therefore bound up with the peculiar state and role of theory in SS – i.e., lack of theoretical models and employment of inter- and trans-disciplinary methodologies and models.

4.3 Purposes of sustainability experiments

Next to declaring that normative knowledge (i.e., evidence on different sustainability ideals and how different agents prioritize them in a specific context) is essential for sustainability experimentation, few studies have examined this problem in depth. However, outputs, outcomes, functions and purposes of sustainability experiments have recently become a focus of research. For example, Kivimaa et al. (2017a, b) problematized actual goals and functions of experiments by distinguishing the various purposes experiments may have (e.g. niche creation, market creation, societal problem solving, etc.). According to these authors, analyzing outputs and outcomes in relation to the purpose, actual goals and wider objectives of the experiment are important determinants in the classification of experiments, and are critical for sound ex-post evaluation of evidence. Laakso et al. (2017) developed this line of thinking further by classifying experiments according to four functions they can serve in sustainability transformations: testing, influencing, multiplying influence, and eventually, promoting systemic change. Both studies point out that since experiments are such a major explanatory concept, sustainability scholars must be much more specific about their nature, characteristics and, particularly, the purposes they are used for and their expected outcomes (Kivimaa et al., 2017a, b; Laakso et al., 2017).

Following the same reasoning, Sengers et al. (2019) focuses on normative orientation of experiments, understood as the changes the experiments aim at (e.g., transition experiments, niche experiments, grassroots experiments, etc.). But are the experimenters’ options for normative orientation limited to these broad types? Or are there different normative visions and value assumptions cutting across the general types of attempted societal change? Are the normative orientations belonging to

the same type (e.g., niche experiments) always comparable? And does belonging to the same type enable transferability of experimental results?

What is strikingly missing from the above literature is a discussion of research value assumptions, which might influence how the research problem is framed, as well as what counts as a solution. For example, while the goal of niche experiments is to help achieve global environmental sustainability by creating or strengthening green technological niches, this can be approached with a variety of research-assumptions and criteria, including researchers' norms influencing which factors should be included and which should not, or what the indicators and thresholds of sustainability are. Therefore, the standards and norms against which experiments are (or should be) evaluated remain obscure, and with them the barriers to external validity.

The overlapping taxonomies discussed in this section illustrate how different features could be used to identify and classify sustainability experiments. Next to the above discussed dimensions, the key features one might take include the conceptual underpinnings of different experiments (Sengers et al., 2019), the methods and instruments used (e.g. from quantitative to qualitative); the types of interventions and knowledge aims (Caniglia et al., 2017); the outputs, functions and purposes (Kivimaa et al., 2017a, b; Laakso et al., 2017); roles in the governance of sustainability (Bulkeley and Cast an Broto, 2013); the types of experimental context and setting (Karvonen & van Heur, 2014; Schapke et al., 2018); or types of researchers' enrollment (as observers or variously engaged participants) (Weiland et al., 2017). To acknowledge this pluralism of perspectives, it is important both to understand any taxonomy as essentially being vague and open-ended, and to perceive the suggested types of experimentation as simply illustrations, categorized according to specific analytical purposes.

Building on the above theoretical perspectives, the next section will focus on the normative orientation of experiments, as particularly relevant for understanding and optimizing evidence traveling. What I will argue below is that experiments in the same category of functions or purposes of experimentation can nevertheless have a different normative orientation in the sense of essentially different sustainability visions, epistemic and ethical assumptions, the sustainability values tested and promoted, and therefore experimental standards of success, which may prevent their comparison and sharing of the results.

5 Discussion: normativity and complexities of evidence traveling in sustainability science

The main questions I am interested in in this section concern: 1) *structural relationships between experiments and the evidence they produce*; and 2) *what the prospects for enhancing transferability and generalizability of the results are* (i.e. evidence traveling). To tackle this topic, it is important to have in mind that evidence is a highly contested notion in SS, as well as in science in general (Douglas, 2000; Kincaid et al., 2007), due to the pervasive roles of values in research. This is commonly known in philosophy in the form of argument on inductive risk

in hypothesis testing (claiming that ethical values affect the choice of methodology and acceptance of hypothesis). Although it has some limitations (e.g. Bidle, 2016), the argument is particularly important in the philosophy of SS because of the inherently normative character of the field emphasized by its solution-oriented approach (Miller et al., 2014). Because part of the experimental context in SS is determined by intricate normative sustainability visions (i.e., value-laden target and transformation knowledge), experimental results may depend heavily on the normative context shared by the experimenters (their implicit and explicit value assumptions).⁷ A part of this normative context may consist of tacit knowledge of the experimenters or symbolic knowledge of the participants and it is strongly tied to the habits and norms of social (including scientific) groups (Asheim et al., 2007). This may be particularly difficult to identify, understand and relate to other experiments. Due to these normative complexities, as in the other social science fields, when evaluating experiments in SS one faces the need to have a deeper interpretative understanding of the experimental context, adding a layer of complexity to transferability issues.

Sustainability-specific value issues place important limitations on the transferability and generalizability of results in SS. First, a sufficiently clear understanding of the context and some level of consensus regarding it is required from experiment evaluators (requiring close study of the values and principles at play, which are often vague and implicit). As researchers' values affect the problem and solution framing (Douglas, 2000), the output of experimentation depends on normative assumptions underlying the research. Second, transfer of evidence is justified (externally valid) only across cases with sufficiently similar normative contexts, specifically the participants' values regarding sustainability targets and transformation pathways.⁸ To illustrate the idea of 'sufficiently similar normative contexts', compare, for example, experiments for sustainable food systems aimed at increasing the efficiency of the current system (e.g. sustainable intensification) relative to those aiming for fundamental re-design of the system (e.g. agro-ecology) (Clapp, 2018). The highly contested debate between these two food system approaches is a radical example of diverging scientific approaches resulting from mutually incompatible systems of research value assumptions (Stojanovic, 2020). Although both target sustainability in the agri-food sector, evidence resulting from one approach is usually irrelevant or regarded as neither internally nor externally valid from the perspective of the other approach (even for experiments having the same function or purpose, e.g. niche experiments), because the basic normative assumptions are considered (theoretically or practically) irrelevant or inadequate for sustainability as envisioned from each approach. Stark differences in these two approaches to sustainable food systems are arguably typical for the broadest approaches to sustainability (although the

⁷ For an exhaustive list of other important contextual factors relevant for sustainability experiments, next to normativity, see Heiligenberg et al. (2017).

⁸ The argument in this paper is focused on target and transformation knowledge, characteristic of sustainability transformation experiments. However, the argument from inductive risk (Douglas, 2000) undermines also the causal, system knowledge. As the latter argument is part of a general discussion on values in science (see Kincaid et al., 2007) which is above the scope of this paper, we focus this section on evidence traveling in solution-oriented experiments.

problem pertains even to technical experiments (Bilali, 2019)). For example, consider radically different sets of values behind the sustainable development agenda (UN, 2015) on one hand, and on the other, degrowth or ecological economics approaches (Bonnedahl & Heikkurinen, 2018). My suggestion is that this correlation in the details of normative orientation matters for evidence production and traveling (despite the shared adherence to general sustainability). As the food systems example illustrates, the relevance and transfer of experimental evidence increases when studies share their value assumptions. This is an under-researched topic in the philosophy of scientific experimentation, but interestingly, we can observe similar phenomena in many practitioner-led experiments, like urban food sustainability or low-carbon community initiatives, in which the entire (sub)fields are based on community experiments and pilot projects sharing the same (and relatively specific) normative assumptions, goals and visions (e.g. Chance et al., 2018).

In a recent literature review on sustainability, Moore et al. (2017) found that 185 of the 209 articles reviewed (88.5%) did not include a definition of ‘sustainability’ in their research, not even a general one. In practice, and often implicitly, a general definition of sustainability (usually the one provided by the Brundtland Report (Brundtland, 1987)) is ordinarily conflated with aiming to increase the efficiency of the social or technological activity – which many argue is actually a contradictory understanding of sustainability. (For a recent forceful defense of this position, see Bonnedahl & Heikkurinen, 2018; also Schapke & Rauschmayer, 2014). In any case, much more nuanced and explicit understanding of intricate value commitments of SS are necessary, particularly in the context of the UN Agenda 2030 (UN, 2015) (Schneider et al., 2019).

Focusing on the normativity inherent in sustainability transformation experiments, the same problem prevails because the main transformation theories (transition management and reflexive governance) do not explicitly address the normative dimension of sustainability transformations (Weiland et al., 2017). Although both theories take a functionalist perspective in the sense of studying options ‘that work’ in the promotion of change towards sustainability, the target of the transformation typically remains vague and general, and it is somehow expected that more sustainable options automatically emerge from experimentation. As Weiland et al. noted: “*The question, however, is how to ensure that the transformation is going in the ‘right’ direction.*“ (2017). The key for this, as I argued, is to challenge idealistic, objectivistic ideas on experimentation as the production of facts (Kincaid et al., 2007), and systematically address the values underlying the research.

On the research-framing side, although sustainability cannot be translated into a completely defined end state, some deliberation on the assumptions and sharing them is necessary (basic structural similarity of the normative dimensions), if experiments are going to belong to the same evidence sharing class, instead of assuming that different resulting normative orientations will always be comparable in terms of evidence traveling. However, as participants are often experiment designers and governors (see §3.2), researchers must explicitly identify both their own value-laden assumptions and participants’ values, and then investigate how these values interact with production of facts (cf. Potthast, 2015). In other words, to produce the desired type of evidence, research has to consider basic normative issues (many of them

already part and parcel of the social sciences) such as ‘evidence *of* what (which normative system)?’ and ‘evidence *for* what (which normative purpose)?’. In SS this particularly becomes, ‘what system we want to sustain’ and ‘for which and whose purpose?’ (Piso, 2016). In current experimental research practice, many of these questions are tackled only at a general level, with detailed discussions of implicit research commitments avoided due to positivistic biases (the ideal of value free science), or for reasons of unclarity in the UN’s Agenda 2030, stakeholder inclusiveness and other ethical or political worries (Schneider et al., 2019).

Because the existing taxonomies (§3) omit both the contextual ties of experiments (notably value preferences of the experimental subjects) and value assumptions of the researchers, they do not adequately capture the normative dimension of sustainability experiments, thereby obscuring the barriers for transferability of results. As both are connected to interpretative contextual information, some have argued that bringing the natural and social sciences together is a necessary precondition for the success of SS (Kates et al., 2001; Jerneck et al., 2011). This applies to general methodological differences, such as those between the goals of interpretation guiding some social sciences and the goals of explanation and prediction guiding natural science. Without taking normative issues seriously (traditionally tackled with interpretative methodologies), it is hard to pinpoint which class of objects experimental results pertain to – what explains the observed low transfer of evidence in SS (Adler et al., 2018). To facilitate this deliberation, we first need more studies mapping the sustainability values at play, both in society (Plieninger et al., 2013) and in sustainability-research (Stojanovic, 2020). Only then we can identify the key values (cf. Horcea-Milcu et al., 2019) and how they change in the course of research and assess the normative landscape of SS and its potential for evidence traveling.

An additional suggestion to grasping the transferability problem, particularly in sustainability transition experiments, is improving the performance of experiments through societal learning frameworks (Luederitz et al., 2016). Considering that many sustainability experiments strive to provide learning opportunities (for both subsequent experiments and for stakeholders and participants), there is a growing realization in the experimental literature of the crucial importance of societal learning from multiple experiments (McFadgen & Huitema, 2018; Luederitz et al., 2016). Here it is important to broaden the collection of relevant data and include at least some core social values in this analysis, considering, for example, altruistic motives (Schapke & Rauschmayer, 2014) or social norms for cooperation (Davis et al., 2018) in behavioral models for sustainability transitions as one of the factors relevant for enhancing societal learning. Another variable that social learning frameworks must include is acknowledgment of different kinds of knowledge, such as indigenous and local knowledge (Tengö et al., 2017), which may contain important insights for classifying relevant scientific evidence (implying also more attention to multi-generational sustainability-related natural experiments). Currently, as we still lack extensive mapping of sustainability-related normative social mechanisms, particularly regarding value creation and value change processes, we need a more integrated system-level perspective (Ison and Straw, 2020) for the analysis to go forward. Despite some early analyses spawned recently in the context of value co-creation literature, as a scientific community, SS is just beginning to grasp the shape of this crucial

topic and potential of sustainability experimentation for robust and usable evidence generation.

In summary, although the available categorizations of experiments are useful for capturing the current diversity of sustainability experimentation and also for contrasting it with classical experimentation, interpretative contextual analysis (notably of the normative dimension) is required to have a systematic understanding of evidence production and traveling in SS. The plethora of purposes and sustainability assumptions and goals present in sustainability research stand in a need of a deeper systemic analysis to overcome research limitations coming from the vague commitments to generalized sustainability. Currently, even reducing greenhouse gas emissions is ordinarily approached in a non-comprehensive way and the mixing of ‘reducing overall emissions from an activity’ with ‘increasing the efficiency of a specific process’ abounds in the SS literature. This conceptual mess regarding what I call *normative mismatch* among different experiments at least partially explains the lack of transferability of results. This is an under-researched topic, and my suggestion would be to look in more detail at how current evidence traveling among the studies creates clustering in the sustainability research along the lines of normative orientation. Based on what we can already observe in studies about sustainable food systems, development economics, etc., I find that normative alignment and evidence sharing are correlated in SS and drive the field in a cluster-like direction, forming subfields for evidence sharing (and methodologically alienating the approaches – dynamics which may eventually endanger the unity of SS as a field).

6 Concluding remarks

The aim of the present study was to contribute to a deeper understanding of sustainability experimentation in the context of the philosophy of experimentation. First, it was investigated how key features of sustainability experimentation evolved from different experimental fields, notably biology and the social sciences, and which were the main driving forces of this development from a philosophical perspective. Then main taxonomies of sustainability experimentation were studied, their motives and drawbacks, with particular focus on the normativity ingrained in sustainability experimentation, how it was covered in the available early attempts to systematize this young field, and what can be done for further conceptual ordering of the assumptions, outcomes and purposes of experimentation. Finally, I tried to bring more depth into the discussion on the normative orientation of experiments and to explain the observed limited traveling of results – two topics that are mostly neglected in the sustainability experimentation literature.

Sustainability experimentation is a novel scientific field, and this paper is one of the early attempts at understanding it better from a philosophy of science perspective. In the paper I argued how the general philosophy of experimentation, together with the philosophies of biology and of the social sciences provide the basic framework for understanding sustainability experimentation. The philosophy of biology provides the basis for the taxonomy of sustainability experimentation and explanation of the attempts to strengthen the relationship between experimental models

and the world. The philosophy of the social sciences introduces the analysis of normativity, context dependency and necessity of interpretative methodologies in the complex meaning- and history-laden object of research. Finally, the general philosophy of experimentation delineates motives for liberalizing experimental control and theoretically grounds the pluralism of the experimental modes characteristic of SS. The analysis was built on these three groups of insights and has integrated them into a comprehensive philosophical picture of sustainability experimentation, orbiting the pluralistic, normative concept of participatory, real-world sustainability experiments.

However, the problem of evidence traveling in SS persists as deeply problematic. I argued that although normativity remains a central problem, some clarity regarding normative barriers and enablers of experimentation can be established with a functionalist approach to sustainability experimentation. The focus here was on the normative orientation of experiments, and how it facilitates transferability across classes of experiments with similarly-defined orientations – epistemic and ethical value assumptions, and purposes of experiments. I claimed that for a more detailed identification of normative orientation of the experiments, more mapping is needed regarding existing values in both science (particularly implicit research assumptions) and in society (research participants). Also, more studies are needed on how values change in the course of experimentation. Engaging in these topics, this research is one of the first philosophical accounts of sustainability experimentation, and it contributes to the further development of an evidence-based SS through gaining a better understanding how evidence is produced in sustainability experiments, how it travels, and what the barriers are to more relevant and usable knowledge.

Funding Open access funding provided by University of Helsinki including Helsinki University Central Hospital. No funding has been received for conducting this research.

Declarations

Ethical approval The manuscript does not require ethical approval.

Informed consent No informed consent was required for conducting this research.

Conflict of Interest The author declares no conflict of interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abson, D. J., Fischer, J., Leventon, J., Newig, J., Schomerus, T., Vilsmaier, U., von Wehrden, H., Abernethy, P., Ives, C. D., Jäger, N. W., & Lang, D. J. (2016). Leverage points for sustainability transformation. *Ambio*, *46*(1), 30–39.
- Adler, C., Hirsch Hadorn, G., Breu, T., et al. (2018). Conceptualizing the transfer of knowledge across cases in transdisciplinary research. *Sustainability Science*, *13*, 179–190. <https://doi.org/10.1007/s11625-017-0444-2>
- Ansell, C. K., & Bartenberger, M. (2016). Varieties of experimentalism. *Ecological Economics*, *130*. <https://doi.org/10.2139/ssrn.2475844>
- Asheim, B. T., Coenen, L., Moodysson, J., & Vang, J. (2007). Regional innovation policy. *International Journal of Entrepreneurship and Innovation Management*, *7*(2–5), 140–155.
- Benz, M. & Meier, S. (2006). Do people behave in experiments as in the field? Evidence from donations. Institute for Empirical Research in Economics Working Paper No. 248.
- Biddle, J. B. (2016). Inductive Risk, Epistemic Risk, and Overdiagnosis of Disease. *Perspectives on Science*, *24*(2), 192–205.
- Bilali, H. E. (2019). Research on agro-food sustainability transitions: A systematic review of research themes and an analysis of research gaps. *Journal of Cleaner Production*, *221*, 353–364. <https://doi.org/10.1016/j.jclepro.2019.02.232>
- Bonnedahl, K. J., & Heikkurinen, P. (Eds.). (2018). *Strongly Sustainable Societies*. Routledge.
- Brandon, R. N. (1996). *Concepts and Methods in Evolutionary Biology*. Cambridge University Press.
- Brown, R. (1997). Artificial experiments on society: Comte, C.G. Lewis and Mill. *Journal of Historical Sociology*, *10*(1), 74–97.
- Brundtland, G. (1987). Report of the World Commission on Environment and Development: Our Common Future. United Nations General Assembly document A/42/427. Oxford University Press
- Bulkeley, H., & Cast an Broto, V. (2013). Government by experiment? Global cities and the governing of climate change. *Transactions of the Institute of British Geographers*, *38*, 361e375. <https://doi.org/10.1111/j.1475-5661.2012.00535>
- Caniglia, G., Schöpke, N., Lang, D. J., Abson, D. J., Luederitz, C., Wiek, A., Laubichler, M., Gralla, F., & von Wehrden, H. (Accepted/In press). Experiments and evidence in sustainability science: A typology. *Journal of Cleaner Production*. <https://doi.org/10.1016/j.jclepro.2017.05.164>
- Chance, E., Ashton, W., Pereira, J., et al. (2018). The Plant—An experiment in urban food sustainability. *Environmental Progress & Sustainable Energy*, *37*(1), 82–90. <https://doi.org/10.1002/ep.12712>
- Clapp, J. (2018). Mega-mergers on the menu: corporate concentration and the politics of sustainability in the global food system. *Global Environmental Politics*, *18*(2), 12–33.
- Clark, W. C. (2007). Sustainability science: A room of its own. *Proceedings of the National Academy Sciences of the United States of America*, *104*(6), 1737–1738. <https://doi.org/10.1073/pnas.0611291104>
- Clark, W. C., van Kerkhoff, L., Lebel, L., & Gallopin, G. C. (2016). Crafting usable knowledge for sustainable development. *Proceedings of the National Academy of Sciences*, *113*(17), 4570–4578.
- Cleland, C. (2001). Historical science, experimental science, and the scientific method. Department of Philosophy and Center for Astrobiology, University of Colorado, Boulder, Colorado 80309, USA. Available at <http://ecee.colorado.edu/ecen5009/Resources/Cleland01.pdf>. Accessed on 4. December 2019.
- Cartwright, N. (1999). *The dappled world*. Cambridge University Press.
- Davis, T., Hennes, E. P., & Raymond, L. (2018). Cultural evolution of normative motivations for sustainable behaviour. *Nature Sustainability*, *1*(5), 218–224.
- Diamond, J. M. (1986). Overview: laboratory experiments, field experiments, and natural experiments. In J. Diamond & T. J. Case (Eds.), *Community Ecology* (pp. 3–22). Harper and Row.
- Dipboye, R. L., & Flanagan, M. F. (1979). Research Settings in Industrial and Organizational Psychology: Are Findings in the Field More Generalizable than the Laboratory. *American Psychologist*, *34*(2), 141–150. <https://doi.org/10.1037/0003-066x.34.2.141>
- Douglas, H. (2000). “Inductive Risk and Values in Science.” Gordon, H. Scott. 2002. *The History and Philosophy of Social Science*. Routledge, second edition. *Philosophy of Science*, *67*, 559–579
- Feynman, R. P., Leighton, R. B., & Sands, M. (1963). *The Feynman Lectures on Physics*. Addison-Wesley.
- Gerber, A., & Green, D. (2011). Field Experiments and Natural Experiments. in Robert E. Goodin (ed.). *The Oxford Handbook of Political Science*. Oxford University Press. Retrieved 19 Jul. 2019, from <https://www.oxfordhandbooks.com/view/>

- Government Programme. (2015). Valtioneuvosto. Accessed August 14, 2019. <http://valtioneuvosto.fi/en/sipila/government-programme>
- Graeber, D. (2001). *Toward an anthropological theory of value*. Palgrave
- Grin, J., Rotmans, J., & Schot, J. (2010). *Transitions to sustainable development: new directions in the study of long term transformative change*. Routledge.
- Grodwohl, J. B., Porto, F., & El-Hani, C. N. (2018) The instability of field experiments: building an experimental research tradition on the rocky seashores (1950–1985). *History and Philosophy of the Life Sciences*, 40(3), 45. <https://doi.org/10.1007/s40656-018-0209-y>
- Grunwald, A. (2004). Strategic knowledge for sustainable development: the need for reflexivity and learning at the interface between science and society. *International Journal of Foresight and Innovation Policy*, 1(1–2), 150–167.
- Guggenheim, M. (2012). Laboratizing and de-laboratizing the world: changing sociological concepts for places of knowledge production. *History of the Human Sciences*, 25(1), 99–118.
- Hacking, I. (1983). *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science*. Cambridge University Press.
- Hausman, J., & Wise, D. (Eds.) (1985). *Social experimentation*. University of Chicago Press for National Bureau of Economic Research.
- Heiligenberg, Harm A. R. M., van den, Gaston J., Heimeriks, Marko P., & Hekkert, Frank G., van. (2017). A habitat for sustainability experiments: Success factors for innovations in their local and regional contexts. *Journal of Cleaner Production*, 169, 204–215.
- Hilden, M., Jordan, A., & Huitema, D. (2017). Special issue on experimentation for climate change solutions editorial: The search for climate change and sustainability solutions - The promise and the pitfalls of experimentation. *Journal of Cleaner Production*, 169, 1–7.
- Horcea-Milcu, A., Abson, D. J., Apetrei, C. I., et al. (2019). Values in transformational sustainability science: four perspectives for change. *Sustainability Science*, 14, 1425–1437. <https://doi.org/10.1007/s11625-019-00656-1>
- Jalas, M., Hyysalo, S., Heiskanen, E., Lovio, R., Nissinen, A., Mattinen, M., Rinkinen, J., Juntunen, J. K., Tainio, P., & Nissilä, H. (2017). Everyday experimentation in energy transition: a practice-theoretical view. *Journal of Cleaner Production*. <https://doi.org/10.1016/j.jclepro.2017.03.034>
- Jerneck, A., Olsson, L., Ness, B., Anderberg, S., Baier, M., Clark, E., Hickler, T., Hornborg, A., Kronsell, A., Lövbrand, E., & Persson, J. (2011). Structuring sustainability science. *Sustainability Science*, 6(1), 69–82.
- Karvonen, A., & van Heur, B. (2014). Urban laboratories: experiments in reworking cities. *International Journal of Urban and Regional Research*, 38(2), 379–392.
- Kates, R. W., Clark, W. C., Corell, R., Hall, J. M., Jaeger, C. C., Lowe, I., McCarthy, J. J., Schellnhuber, H. J., Bolin, B., Dickson, N.M., & Faucheux, S. (2001). Sustainability science. *Science*, 292(5517), 641–642.
- Kincaid, H., Dupré, J., & Wylie, A. (Eds.). (2007). *Value-free science: ideals and illusions?*. Oxford University Press.
- Kitcher, P. (2019). So ... who is your audience? *European Journal for Philosophy of Science*, 9, 2. <https://doi.org/10.1007/s13194-018-0227-3>
- Kivimaa, P., Hilden, M., Huitema, D., Jordan, A., & Newig, J. (2017a). Experiments in climate governance e a systematic review of research on energy and built environment transitions. *Journal of Cleaner Production*. <https://doi.org/10.1016/j.jclepro.2017.01.027>
- Kivimaa, P., Hildén, M., Huitema, D., Jordan, A., & Newig, J. (2017b). Experiments in climate governance – A systematic review of research on energy and built environment transitions. *Journal of Cleaner Production*, 169, 17–29. <https://doi.org/10.1016/j.jclepro.2017.01.027>
- Krohn, W., & Weyer, J. (1994). Society as a laboratory: The social risks of experimental research. *Science and Public Policy*, 21(3), 173–183.
- Laakso, S., Berg, A., & Annala, M. (2017). Dynamics of experimental governance: A meta-study of functions and uses of climate governance experiments. *Journal of Cleaner Production*, 169, 8–16. <https://doi.org/10.1016/j.jclepro.2017.04.140>
- Lang, D. J., Wiek, A., Bergmann, M., et al. (2012). Transdisciplinary research in sustainability science: practice, principles, and challenges. *Sustainability Science*, 7(1), 25–43.
- Latour, B. (2011). From multiculturalism to multinaturalism: what rules of method for the new socio-scientific experiments? *Nature and Culture*, 6(1), 1–17.
- Levitt, S., & List, J. (2008). Field experiments in economics: The past, the present, and the future. *European Economic Review*, 53(1), 1–18.

- Luederitz, C., Schapke, N., Wiek, A., Lang, D. J., et al. (2016). Learning through evaluation: a tentative evaluative scheme for sustainability transition experiments. *Journal of Cleaner Production*, 169, 61–76.
- MacLeod, M., & Nagatsu, M. (2018). What does interdisciplinarity look like in practice: Mapping interdisciplinarity and its limits in the environmental sciences. *Studies in History and Philosophy of Science Part A*, 67, 74–84.
- Mandler, G. (2007). *A history of modern experimental psychology: From James and Wundt to cognitive science*. MIT Press.
- McFadden, B., & Huitema, D. (2018). Experimentation at the interface of science and policy: A multi-case analysis of how policy experiments influence political decision-makers. *Policy Science*, 51, 161–187. <https://doi.org/10.1007/s11077-017-9276-2>
- Miller, T. R., Wiek, A., Sarewitz, D., Robinson, J., Olsson, L., Kriebel, D., & Loorbach, D. (2014). The future of sustainability science: A solutions-oriented research agenda. *Sustainability Science*, 9(2), 239–246.
- Mitchell, S. D. (2009). *Unsimple Truths: Science, Complexity, and Policy*. University of Chicago Press.
- Moore, J. E., Mascarenhas, A., Bain, J., et al. (2017). Developing a comprehensive definition of sustainability. *Implementation Science*, 12, 110. <https://doi.org/10.1186/s13012-017-0637-1>
- Nagatsu, M., Davis, T., DesRoches, C. T., et al. (2020). Philosophy of science for sustainability science. *Sustainability Science*, 15, 1807–1817. <https://doi.org/10.1007/s11625-020-00832-8>
- Nowotny, H. (2015). *The Cunning of Uncertainty*. John Wiley & Sons.
- Piso, Z., et al. (2016). Sustainability of What? Recognising the Diverse Values That Sustainable Agriculture Works to Sustain. *Environmental Values*, 25(2), 195–214.
- Plieninger, T., Dijks, S., Oteros-Rozas, E., & Bieling, C. (2013). Assessing, mapping, and quantifying cultural ecosystem services at community level. United Nations Sustainable Development Platform (2015). Transforming Our World: The 2030 Agenda for Sustainable Development. *Land Use Policy*, 33, 118–129.
- Potthast, T. (2015). Ethics in the sciences beyond Hume, Moore and Weber: taking epistemic-moral hybrids seriously. In *Ethics of science in the research for sustainable development* (pp. 129–152). Nomos Verlagsgesellschaft mbH & Co. KG.
- ProClim, C. A. S. S. (1997). Research on sustainability and global change—visions in science policy by swiss researchers.
- Radder, H. (ed.) (2003). *The Philosophy of Scientific Experimentation*. University of Pittsburgh Press.
- Ison, R., & Straw, E. (2020). *The Hidden Power of Systems Thinking: Governance in a Climate Emergency*. Routledge.
- Rotmans, J., Kemp, R., & van Asselt, M. (2001). More evolution than revolution: transition management in public policy. *Foresight*, 3(1), 15–31.
- Schapke, N., & Rauschmayer, F. (2014). Going beyond efficiency: including altruistic motives in behavioral models for sustainability transitions to address sufficiency. *Sustainability: Science, Practice and Policy*, 10(1), 29–44.
- Schapke, N., et al. (2018). Jointly Experimenting for Transformation? Shaping Real-World Laboratories by Comparing Them. *GAIA - Ecological Perspectives for Science and Society*, 27(S1), 85–96.
- Schneider, F., Kläy, A., Zimmermann, A. B., et al. (2019). How can science support the 2030 Agenda for Sustainable Development? Four tasks to tackle the normative dimension of sustainability. *Sustainability Science*, 14, 1593–1604. <https://doi.org/10.1007/s11625-019-00675-y>
- Schot, J., & Geels, F. W. (2008). Strategic niche management and sustainable innovation journeys: theory, findings, research agenda, and policy. *Technology Analysis & Strategic Management*, 20(5), 537–554.
- Sengers, F., Wiczorek, A. J., & Raven, R. (2019). Experimenting for sustainability transitions: a systematic literature review. *Technological Forecasting and Social Change*. <https://doi.org/10.1016/j.techfore.2016.08.031>
- Shove, E., & Walker, G. (2010). Governing transitions in the sustainability of everyday life. *Research Policy*, 39, 471–476.
- Stojanovic, M. (2013). Exploratory Experimentation and Taxonomy of Experimentation. *Philosophy and Society Xxiv*, 4, 199–217. <https://doi.org/10.2298/FID1304199S>
- Stojanovic, M. (2020). Divergent approaches to sustainable food systems research_ Value assumptions as a source of controversy. Forthcoming. Manuscript presented at the conference “Interdisciplinarity, Sustainability and Expert Disagreement _ Philosophical bias in science, motivating scientific

- controversy over sustainability". Oslo, 10 – 11 June 2020, Norwegian University of Life Sciences (NMBU).
- Tengö, M., Hill, R., Malmer, P., Raymond, C. M., Spierenburg, M., Danielsen, F., Elmqvist, T., & Folke, C. (2017). Weaving knowledge systems in IPBES, CBD and beyond—lessons learned for sustainability. *Current Opinion in Environmental Sustainability*, 26–27, 17–25. <https://doi.org/10.1016/j.cosust.2016.12.005>
- Thye, S. R. (2014). Logical and philosophical foundations of experimental research in the social sciences. In *Laboratory experiments in the social sciences* (Second Edition), Chapter 3, pp. 53–82. <https://doi.org/10.1016/B978-0-12-404681-8.00003-0>
- UN. (2015). United Nations resolution. *Transforming Our World: The 2030 Agenda for Sustainable Development*. https://www.un.org/ga/search/view_doc.asp?symbol=A/RES/70/1&Lang=E
- VanderVeen, Z. (2011). John Dewey's experimental politics: inquiry and legitimacy. *Transactions of the Charles S. Peirce Society*, 47, 158. <https://doi.org/10.2979/trancharpeirsoc.47.2.158>
- Weiland, S., Bleicher, A., Polzin, C., Rauschmayer, F., & Rode, J. (2017). The nature of experiments for sustainability transformations: A search for common ground. *Journal of Cleaner Production*, 169, 30–38. <https://doi.org/10.1016/j.jclepro.2017.06.182>
- Wiek, A., Ness, B., Schweizer-Ries, P., Brand, F. S., & Farioli, F. (2012). From complex systems analysis to transformational change: a comparative appraisal of sustainability science projects. *Sustainability Science*, 7, 5–24.
- Wilson, B. (2009). *Nature as Laboratory: Experiments in Ecology and Evolutionary Biology*. Presented at the Second Biennial Conference of the Society for the Philosophy of Science in Practice, June 18–20, 2009, Minneapolis, Minnesota.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.