UNIVERSITY OF LEEDS

Evidence synthesis methods to enhance climate policy evaluation for IPCC assessments

Niklas Döbbeling-Hildebrandt

Submitted in accordance with the requirements for the degree of

Doctor of Philosophy

The University of Leeds

School of Earth and Environment

December 2024

Declaration of authorship

I confirm that the work submitted is my own, except where work which has formed part of jointly authored publications has been included. My contribution and the other authors to this work has been explicitly indicated below. I confirm that appropriate credit has been given within the thesis where reference has been made to the work of others.

Chapter 2 has been co-authored with Tim Repke, Arianna Avallone, Finlay Hatch, Alessandra Landa, Klaas Miersch, Leonhard Schneider, and Jan C. Minx. It will be submitted for publication after the submission of this thesis.

I designed the research, with contributions from T.R. and J.C.M.. I developed the literature screening strategy, with contributions from T.R.. I, A.A., F.H., A.L., K.M., and L.S. manually screened the literature and extracted the data. I performed the machine learning-enabled screening together with T.R.. T.R. performed the machine-learning predicted clustering, in close collaboration with me. T.R. developed an interactive online repository for the data. I analysed the results and discussed the findings with T.R. and J.C.M.. I wrote the manuscript, with contributions and edits from T.R. and J.C.M..

Chapter 3 has been co-authored with Klaas Miersch, Tarun M. Khanna, Marion Bachelet, Stephan B. Bruns, Max Callaghan, Ottmar Edenhofer, Christian Flachsland, Piers M. Forster, Matthias Kalkuhl, Nicolas Koch, William F. Lamb, Nils Ohlendorf, Jan Christoph Steckel, and Jan C. Minx. It was published in Nature Communications as *Systematic review and meta-analysis of ex-post evaluations on the effectiveness of carbon pricing*, in 2024 (DOI: 10.1038/s41467-024-48512-w).

I designed the research, with contributions from J.C.M., K.M., T.M.K., O.E., N.K., W.F.L., N.O., and J.C.S.. I developed the literature screening strategy with the help from K.M., M.C., and J.C.M.. I, K.M., M.B., N.K., W.F.L., N.O., J.C.S., and J.C.M. manually screened the literature. I, K.M., and M.B. extracted the data. I performed the machine learning-enabled screening together with M.C.. I performed the meta-analysis with the help from K.M., T.M.K., and S.B.B.. Together with T.M.K., I conducted the Bayesian model averaging analysis. I analysed the results and discussed the findings with K.M., T.M.K., S.B.B., O.E., C.F., P.M.F., M.K., N.K., J.C.S., and J.C.M.. I wrote the manuscript, with contributions from T.M.K., J.C.M., and K.M. and edits from all authors.

Chapter 4 has been co-authored with Diana Danilenko, William F. Lamb, and Jan C. Minx. It will be submitted for publication after the submission of this thesis.

I designed the research and the data collection strategy. I collected and synthesised the data. I analysed the results with contributions from W.L, D.D., and J.C.M.. I wrote the manuscript, with edits from D.D., W.L., and J.C.M..

Acknowledgements

I would like to thank the Heinrich Böll Stiftung for the financial support provided for this PhD, and the Mercator Research Institute on Global Commons and Climate Change (MCC), the Priestley Centre for Climate Futures, and the School of Earth and Environment at the University of Leeds for all the support provided throughout my PhD. I heartily thank my supervisors Jan Minx and Piers Forster for their advice and feedback that significantly contributed to the success of this thesis. I thank all my co-authors for their valuable contributions to this work. I am very grateful to my colleagues at MCC for the collaborative and supportive working environment. I also want to thank my partner, my friends, and my family for their great mental support that helped me to complete this thesis.

Abstract

Climate policy research has greatly accelerated in recent years, resulting in the publication of tens of thousands of policy evaluations. With immense efforts, the Intergovernmental Panel on Climate Change (IPCC) assesses this research, aiming to provide policymakers with the best available evidence to assist the choices and designs of climate policies. Yet, in the wider policy evaluation community, evidence synthesis practices, which could support the existing process, are largely missing.

In this thesis, I adopt and refine three evidence synthesis methods from other scientific disciplines to test their applicability and showcase their usefulness for the systematic assessment of the state of knowledge on climate policy questions, specifically on the empirical evidence on carbon pricing. I map out the available evidence on carbon pricing policy evaluations, assessing the state of research and identifying research gaps; conduct a systematic review and meta-analysis on the effectiveness of carbon pricing policies in reducing greenhouse gas emissions; and assess under what conditions and why carbon pricing is or is not effective using a realist synthesis method.

The evidence syntheses conducted here reveal a number of policy relevant research findings that have been missed by previous IPCC reports. The mapping identifies more than 4,000 evaluations of carbon pricing policies, predominantly focused on only few policy schemes, with considerable research gaps for many other schemes. The meta-analysis shows that carbon pricing has effectively reduced greenhouse gas emissions in at least 17 out of 21 reviewed carbon pricing schemes, with reductions of between -5% and -21%. The realist synthesis captures nine causal mechanisms

on how carbon pricing triggers emissions reductions, with varying relevance across sectors and country contexts.

A more widespread application of evidence synthesis methods to address climate policy questions could greatly support the IPCC in providing a meaningful and comprehensive assessment of the climate policy options.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

By signing the Paris Agreement, governments around the world have committed to limit global warming to well below 2°C and to pursue efforts to limit global warming to 1.5°C [1]. Yet, the implementation of climate mitigation policies is not on track to reduce greenhouse gas emissions at a scale required to meet the goals set out by the Paris Agreement [2, 3]. Systematic and rapid climate policy learning is therefore of greatest importance to accelerate the reduction of emissions. Having a sound understanding of what policies work, under what conditions, and why is crucial for progress [4–6].

The good news is that with the experience from thousands of climate policies implemented around the globe [7–11] and considerable scientific efforts, the world has gathered vast knowledge of how policies need to be designed to curb the climate crisis. Policy databases list more than 5,000 climate policies, implemented in 200 countries of the world [12, 13] and recent assessments of the scientific literature find more than 400,000 academic publications studying climate change [14], of which more than 80,000 specifically evaluate mitigation policies [15]. Significant variations in policy instruments and their effectiveness [16–18] provide ideal conditions for cross-country policy learning.

The vast amounts of available evidence, however, are not yet systematically exploited to effectively inform policy decisions. An evidence base that is scattered across large numbers of research articles, published in hundreds of disciplinary journals is practically inaccessible for policymakers and increasingly also for researchers. Even established global environmental assessment bodies, like the Intergovernmental Panel on Climate Change (IPCC), are more and more overburdened by the large amount of available evidence. Global environmental assessments are meant to assess all available evidence to provide a comprehensive account of the science on an environmental issue like climate change. To fulfil this mandate, IPCC assessments have steadily increased the amount of considered evidence throughout its six assessment cycles. Sill, the increase in considered evidence was outpaced by an even faster growth in the available evidence, leading to a decline in relative coverage of the climate change evidence [5].

To keep up with the increase in evidence, assessments of the available research cannot be left to environmental assessment bodies alone, but need to become a collective task in the research community. As traditional literature reviews are increasingly criticised for a variety of biases and their summaries are neither comprehensive nor reproducible [19–21], more rigorous approaches are required to provide (as) unbiased (as possible) accounts of the evidence. Methods for comprehensive, systematic, and transparent assessments of evidence are available in other research fields [22–28]. These need to be adjusted for the specific multidisciplinary context and be systematically applied in the social sciences of climate policy [4, 5].

In this thesis, I test and apply evidence synthesis methods for the evaluation of climate policies. In particular, I create an evidence and gap map of more than 4,000 policy evaluations and conduct two systematic reviews using meta-analysis and realist synthesis approaches, refining the methods to fit the purpose of climate policy research. I evaluate the methods for their feasibility and identify remaining challenges in their application. In Chapter 5, I discuss how global environmental assessments, researchers, and scientific policy advice may benefit from a wider application of evidence synthesis methods to evaluate climate policies.

For the development and testing of the methods in this PhD thesis, I apply the methods to the use case of carbon pricing. This market-based policy measure provides a valuable application, with more than 70 jurisdictions applying a carbon price to date [29], while scientific and political disputes remain on the suitability of the policy measure to mitigate climate change.

1.2 State of evidence synthesis in climate research

1.2.1 Evolution of evidence synthesis across disciplines

About 50 years ago, Gene Glass proposed meta-analysis as a formalised research approach to synthesise research results across large numbers of studies [30, 31] and, together with Mary Smith, conducted its first application to synthesise the findings of more than 300 psychotherapy evaluations [32]. This statistical method to aggregate primary research findings was picked up by numerous reviews in education and psychological research, over the subsequent decade [33, 34]. Medical research – today the forerunner in evidence synthesis – only started adopting these methods in the late 1980s and early 1990s, with an application particularly towards synthesising the findings from randomised controlled trials [21, 35]. These times also coined the term *systematic review*, which more broadly captures all reviews that apply rigorous and transparent evidence synthesis methods.

The uptake of systematic reviews in medicine was driven by challenges, which are comparable to the current situation in climate change research. Health practitioners where facing increasing amounts of available evidence with limited time and resources to gain a systematic overview of the relevant research, when a medical treatment decision had to be taken [36]. Just like policymakers today being overwhelmed by the amount of relevant research for any given climate policy question.

The formalisation of systematic reviews was largely supported by Cochrane, a network founded to promote the application of high-quality systematic reviews [36]. Cochrane shapes the evolution by providing systematic review guidelines, capacity building, methods development, and publishing peer-reviewed systematic reviews [23].

These advancements in systematic review practices, in turn, feed back to the social sciences, where they originated. Fostered by the foundation of the Campbell Collaboration in 2000, systematic reviews on social, psychological, educational, and criminological interventions have been revitalised [37, 38]. Systematic review methods have also spread into the fields of economics [39, 40], public management [41], environmental sciences [42, 43], to name a few of the disciplines relevant for the research on climate mitigation and adaptation. In none of these disciplines, however, have evidence synthesis methods established to the same degree as in health sciences.

1.2.2 Growing diversity in synthesis methods

Beyond the disciplinary diffusion of evidence synthesis, methods have also been diversified over the years. Meta-analysis as a statistical method of combining quantitative evidence has been accompanied by a variety of supplementary evidence synthesis methods for systematically assessing the state of research.

Systematic review methods have evolved to synthesise heterogenous types of evidence and answer diverse types of research questions. Within Cochrane, methods to systematically review qualitative and mixed qualitative and quantitative evidence are developed since the late 1990s [44, 45]. Methods guidelines, today, provide a whole range of available methods to synthesise these types of evidence [46–49]. This also allowed for a broadening of research questions which can be addressed in systematic reviews [47, 50], going beyond the mere hypothesis testing that underlies quantitative meta-analysis.

Kastner *et al.* [27] collected a wide range of synthesis methods available today (listed in Table 1.1), studying research questions on perceptions, blockages, enablers, conceptualisations, etc. The realist synthesis method should be mentioned here as an example of a more configurative research approach, capable of assessing both qualitative and quantitative evidence. The method, which will be used in Chapter 4, was developed by Ray Pawson and colleagues in the early 2000s and aims to address the question under what conditions and why a given intervention works [6, 28].

Systematic review approaches, providing detailed assessments of the evidence on specific research questions, were further supplemented by mapping approaches, aiming to provide broader overviews of a research field [81, 82]. Traditional bibliometric and scientometric methods are heavily focussed on quantifying the literature land-

Synthesis method	Type of reviewed evidence	Examples
Critical interpretative synthesis	qualitative & quantitative	[51, 52]
Integrative review	qualitative & quantitative	[53–55]
Narrative synthesis	qualitative & quantitative	[56–58]
Realist synthesis	qualitative & quantitative	[59–61]
Meta-ethnography	qualitative	[62, 63]
Meta-interpretation	qualitative	[64, 65]
Meta-summary	qualitative & quantitative	[66, 67]
Meta-study	qualitative	[68]
Meta-synthesis	qualitative	[69, 70]
Mixed studies review	qualitative & quantitative	[71]
Meta-narrative review	qualitative & quantitative	[72, 73]
Concept synthesis	qualitative	[69, 74]
Meta-analysis	quantitative	[75–77]
Network meta-analysis	quantitative	[78, 79]

Table 1.1: **Heterogeneity of synthesis methods for systematic reviews**: The table is replicated from Kastner *et al.* [27], who conduct a scoping review of available synthesis methods, with a focus on methods used less frequently than the standard quantitative synthesis methods. These quantitative synthesis methods (meta-analysis and network meta-analysis) are added from Higgins *et al.* [23]. The table is replicated here to provide an idea of the variety of methods, without any claim on completeness. Methods in this field are still evolving and there are other conceptualisations available [48, 49, 80].

scape and citation practices [83, 84]. To provide a more qualitative description of what evidence is out there, social sciences have developed mapping approaches [26, 85]. These approaches, termed *systematic mapping*, *evidence mapping*, or *evidence and gap mapping*, aim to assess what topics have been studied, using what research methods, to provide a topical overview of what research questions have been addressed and where are remaining gaps in the available evidence [81, 82, 86, 87]. Mapping approaches cover a larger breadth of evidence as compared to systematic reviews, with less information extracted from each covered primary study.

More recent advancements of evidence synthesis methods are the introduction of machine-learning approaches and the concept of constantly updated *living* evidence syntheses. Evidence syntheses conduct comprehensive data collections including a range of highly repetitive tasks. Machine-learning approaches are increasingly used for some of these tasks to assist or replace the human reviewer [88–93].

The concept of living evidence was developed over the last decade [94–96] and became popular during the global COVID-19 pandemic, when the need for constantly updated syntheses became ever more apparent [97, 98]. While the frequency of required policy decisions and the amount of published research to address the pandemic was even more extreme than observed for climate change policy, living evidence syntheses could still serve as valuable sources of timely information for policymaking [99].

1.2.3 Global environmental assessments

In parallel to the uptake of evidence synthesis practices described above, the climate research community established a distinct approach for synthesising the available evidence. Global environmental assessments, in particular those conducted by the IPCC, bring together hundreds of field experts to assess the available evidence. Since its foundation in 1988, the IPCC is tasked to provide comprehensive assessments of the climate change research. This involves an intensive interaction between scientists and policymakers, as the scientist's assessment is commissioned, reviewed, and approved by the United Nations member states [100]. In 2023 the seventh assessment cycle commenced, to assess the physical science of climate change, its impacts, as well as options for adaptation and mitigation.

The assessment of climate change mitigation in previous IPCC reports was particularly strong in synthesising the evidence from scenarios, modelling future emissions pathways. For the sixth assessment cycle, the IPCC collected more than 3,000 emissions scenarios [101], conducted quality appraisal, classified the scenarios by their predicted global warming outcomes and by the level of assumed climate action, and synthesised the predicted emissions pathways across models [102]. These efforts build on an active community of researchers conducting model intercomparisons [103, 104].

Similar community efforts have not yet been established for the evaluation of implemented climate policies [4, 5]. Over the past assessment cycles, the IPCC moved from providing more theoretical and ex-ante policy assessments [105, 106] towards more ex-post assessments [2, 107]. Due to a lack of evidence synthesis in this field, the IPCC authors, however, had to draw directly from the large pool of available primary evidence (see Figure 1.1). This challenges the capability of the IPCC to pro-

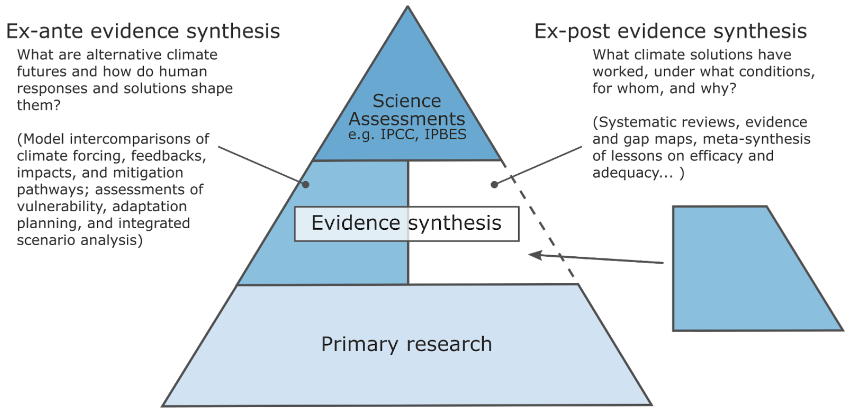

Figure 1.1: **Broken evidence pyramid**: Science assessments require a solid bases of primary evidence and syntheses thereof. The IPCC can build on synthesised ex-ante evidence from model intercomparisons, but lacks a comparable source of synthesised ex-post evidence. This missing link between the large amounts of relevant primary evidence and the high-level assessment, challenges the comprehensive assessment of policy evaluations by the IPCC. The figure was kindly provided by William F. Lamb and is adapted from Berrang–Ford *et al.* [4].

vide a comprehensive and rigorous assessment of all the valuable knowledge gathered by thousands of policy evaluations.

1.2.4 Evidence synthesis in climate research

There is a growing interest within climate related research fields to adopt rigorous evidence synthesis methods to close the gap between primary evidence and high-level science assessments [4, 5, 108]. Yet, evidence synthesis is still sparsely applied in climate research. The availability of machine-learning methods has lead to a number of studies mapping out the literature landscape of different branches of climate change research. With their large scope, they are able to provide helpful orientation within the research field. Systematic reviews are, so far, not very prevalent in the field and, with their lower coverage of the literature, leave wide areas of the climate relevant research unsynthesised. A rapid search for "systematic review" and "meta-analysis" in a dataset of climate policy literature [15] returns fewer than 250 articles.

Maps are a tool that could provide IPCC authors and other users of evidence a comprehensive overview of the available literature. They are designed to capture different aspects of what research has been conducted. Available maps of the climate literature vary in their scope and detail, with larger maps commonly providing fewer information per study than smaller maps with a more specific focus. Another dimension of variation in the published maps is their use of hand-labelling, machine-learning, or their combination.

The largest map is provided by Callaghan *et al.* [14], who aim to map out the entire climate change research landscape. The map is compiled using machine-learning methods and comprises more than 400,000 research articles. These are clustered into 140 research topics using topic modelling. Scientific disciplines are assigned according to the journal of publication. This large mapping effort, unlike other maps in the field, is compiled entirely by machine-learning. It shows the rapid growth in climate change research over the past decades and assesses to what extent different topics have been covered by the IPCC.

More commonly, at least a share of documents is screened and labelled by human coders. Their labelling decisions are then scaled using machine-learning methods to classify unseen documents. Callaghan *et al.* [15] and Sietsma *et al.* [109] use this hybrid approach to map out the climate policy research respectively for mitigation and adaptation, to provide an overview which policies receive higher or lower research attention. Callaghan *et al.* [15] map 84,990 studies on implemented climate change mitigation policies, of which 2,580 are hand-labelled. Sietsma *et al.* [109] classify 8,691 climate change adaptation research papers, of which 2,495 are labelled by hand. Both studies use fine-grained typologies to cluster the studied policy. For mitigation policies the emission sector is labelled, while for the adaptation policies the addressed climate impact is captured. Both studies categorise evidence into ex-post and ex-ante assessments.

A number of studies map the literature on climate change and cities [110–112], with the most recent one finding 55,000 relevant studies (1,020 hand labelled) [113]. The latter study clusters the articles by studied city and applies a topic model to identify research topics. All of the above papers map the literature based on their abstracts.

A more in depth assessment of the literature on climate change adaptation was conducted by the Global Adaptation Mapping Initiative [114], who mapped out 1,682 articles by hand-coding detailed information from the full texts of the studies. It includes a detailed assessment of actors and drivers of adaptation actions, their effectiveness and transformational potential. The map also appraises the study quality of the reviewed literature. The use of machine learning for this study was limited to the screening of literature, where algorithms were applied to prioritise documents based on predicted relevance.

Applications of the evidence and gap mapping approach are conducted for energy efficiency interventions [115] and sustainable energy development [116]. The maps identify 299 and 703 relevant studies, respectively, and categorise them by intervention and outcome. In addition, both maps capture information on the study design and the country where the study was conducted. The latter map also collects some technical information on the energy generation and use.

Systematic reviews in the field of climate research have so far more often reviewed quantitative evidence from ex-ante modelling studies and surveys, rather than from ex-post policy evaluations. A number of systematic reviews also synthesise qualitative evidence.

This thesis is not the first to conduct systematic reviews on carbon pricing policies. Ohlendorf *et al.* [117] review distributional impacts of carbon pricing policies, based on 53 ex-ante modelling studies. The review conducts a meta-analysis including meta-regression to assess the variation in effect sizes. Midões *et al.* [118] review and meta-analyse 54 ex-ante studies on the synergies of carbon pricing with renewable energy policies. There are two systematic reviews on public perception towards carbon pricing [119, 120]. The studies review 35 and 48 survey experiments on the role of revenue recycling for public support. Only Mohammadzadeh Valencia *et al.* [119] conducts a meta-analysis to synthesise the survey findings, while Barrez [120] narratively reviews the primary studies. Interestingly, this year another meta-analysis on the effectiveness of carbon pricing was published, with a similar research question as the systematic review conducted in Chapter 3. This study is further discussed and compared to the findings of my review in section 5.1.2.

Cuevas *et al.* [121] systematically review the available literature on health impacts of carbon pricing. The study synthesises the evidence on health co-benefits and trade-offs from 58 ex-ante modelling studies using a framework analysis. A qualitative content analysis of the ethical arguments for and against carbon pricing is conducted in a systematic review by Magnetti *et al.* [122] reviewing arguments from 210 primary studies.

Beyond the climate change mitigation literature, there are also a number of systematic reviews conducted for climate change impacts and adaptation. Ford *et al.* [123] and Owen [124] conduct content analyses of implemented adaptation measures and their observed effectiveness. Ford *et al.* [123] assess and categorise adaptation measures based on qualitative evidence from 39 primary studies. Owen [124] extracts qualitative and quantitative evidence on the effectiveness of adaptation measures from 110 case studies, synthesising how effectiveness is measured across studies and adaptation initiatives. Méjean *et al.* [125] systematically review the distributional dimension of climate change impacts. The study narratively reviews the quantitative evidence from 127 ex-ante modelling studies.

Climate relevant systematic reviews are also available in the field of behavioural or retrofitting interventions for energy savings. Khanna *et al.* [77] review the effect of behavioural interventions on household energy consumption, using meta-analytic methods to synthesise the research findings from 122 studies. This review is now extended as a living systematic review and will be monthly updated [126]. Willand *et al.* [127, 128] and Camprubí *et al.* [129] systematically review the health outcomes of retrofitting interventions in private households. Both author teams use realist synthesis methods to assess mechanisms and context factors determining the uptake of retrofitting programmes and their impact on health outcomes.

The decades of development and application of evidence syntheses methods in the scientific community, has, until now, only occasionally been picked up by climate relevant research areas. This motivates this thesis to apply various evidence synthesis methods to the field of climate policy assessments and to evaluate their applicability and value for climate policy learning.

1.3 Methods to systematically synthesise evidence

In this thesis, I aim to apply evidence synthesis methods that could help filling the gap between the increasing amount of evidence from ex-post climate policy evaluations and the high-level assessment conducted by the IPCC (see Figure 1.1). I adopt three evidence synthesis methods from across the spectrum of available approaches. The applied methods range from a broad mapping of thousands of ex-post policy evaluations to in-depth reviews synthesising research findings on specific research questions. I explore the spectrum of systematic review methods, conducting an aggregative meta-analysis and a configurative realist synthesis.

1.3.1 Evidence mapping

Evidence mapping is a systematic approach of assessing the literature landscape of a research field. It studies what research is conducted and provides a structured overview of the available evidence [25, 26]. Evidence mapping sets out what evidence should be captured and searches for potentially relevant literature across bibliometric databases. It continues by screening the evidence returned by the literature searches based on clearly defined eligibility criteria. Included studies are commonly categorised based on pre-defined typologies covering topics, methods, and relevant meta-data on the studies. The categories are used to produce visualisations and tables to provide meaningful overviews of the literature. Often, the gathered data is provided in interactive online repositories as a data source for further research [130, 131].

The evidence synthesis method used in Chapter 2 is a so called *evidence and gap map*, characterised by the two main dimensions collected for each study: intervention and outcome. Additional information collected for the included studies can be used for subsampling or filtering the captured literature. The presentation of the literature in a table capturing the two main dimensions provides an insightful overview, which intervention–outcome combinations are studied by how many studies and where gaps in the primary literature remain [26, 86, 130]. This can inform researchers where further primary research is needed or for which research questions the evidence could be synthesised in a systematic review.

I use the method to map out the available evidence from ex-post carbon pricing policy evaluations, categorising the evidence by the studied policies, policy outcomes, emissions sectors, and by the methods used for the evaluation. This provides a detailed overview to inform where the available evidence merits a systematic review and where research gaps should be filled by primary research.

1.3.2 Systematic reviews

Systematic reviews assess the primary literature in more detail, studying a specific research question and synthesising the available evidence to answer the question. A systematic review transparently and comprehensively collects relevant studies, extracts research findings, appraises study designs, and synthesises the evidence from the primary studies [23, 24]. Evaluating the evidence across studies on the same research question can compare and aggregate the findings to provide a more robust answer to the research question. Studying patterns in the reviewed evidence may answer research questions that none of the single studies alone could answer. In this way, systematic reviews make use of the best available evidence to gain new and more robust research insights [23, 24].

While there is some variation in systematic review methods, depending on the research question and type of evidence, there are many communalities across review methods. Similar to mapping approaches, systematic reviews comprehensively search for evidence, retrieving literature across academic and grey literature repositories. A long-list of potentially relevant literature is screened based on transparent criteria for their inclusion or exclusion, while the eligibility criteria are more narrow than in evidence mapping. Once the relevant studies are identified the systematic reviews significantly deviate from mapping approaches. They extract detailed information on the research findings and assess the study designs for potential biases [132, 133]. Data is extracted from the included studies and synthesised to answer the research question of the review. This process has been formalised by several scientific collaborations, like the Campbell Collaboration [22], Cochrane [23], and Collaboration for Environmental Evidence [24], to provide guidance and to ensure high standards for the review. The synthesis of research findings can follow aggregative or configurative approaches, both of which are relevant for this thesis.

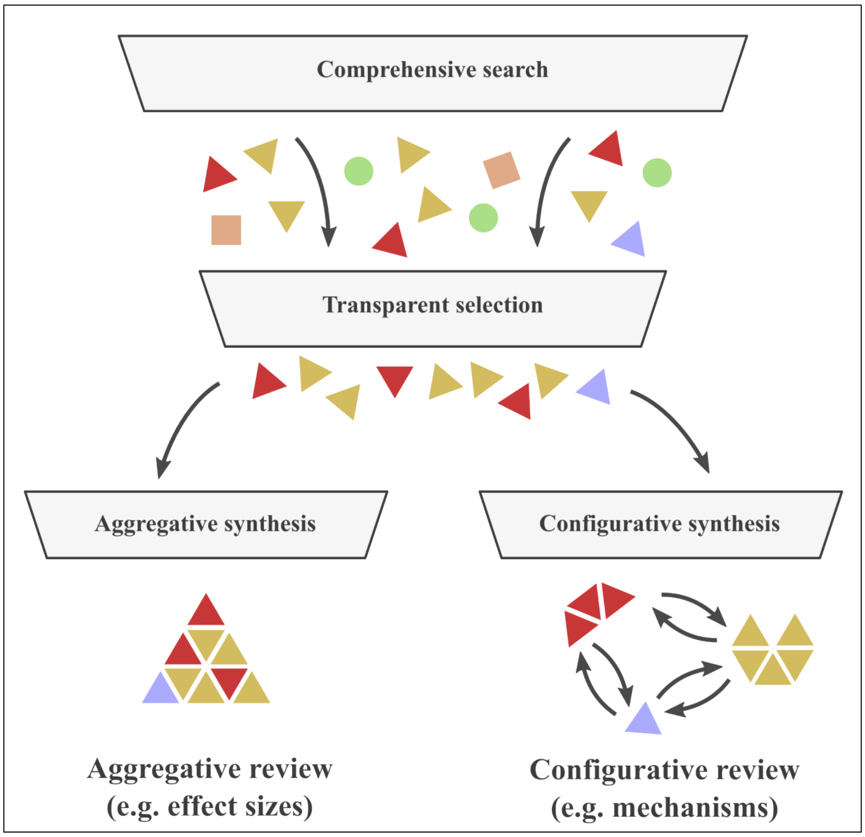

Figure 1.2: **Aggregative and configurative systematic review methods**: Systematic reviews follow a comprehensive and transparent search and selection of relevant evidence. Aggregative reviews use comparable evidence on the same research question to derive a more robust answer. Configurative reviews instead triangulate and interpret evidence to derive new hypotheses and theories.

1.3.3 Aggregative and configurative synthesis

Systematic reviews can be applied to all kinds of research questions, synthesising various types of evidence. Corresponding to the heterogeneity in research questions and types of evidence, there is a wide array of methods available to synthesise the evidence extracted from primary studies (see Table 1.1) [27]. Synthesis methods are often clustered based on the type of evidence to be synthesised, i.e. distinctions are made between qualitative, quantitative, and mixed-methods syntheses [48, 49, 134]. For the two syntheses methods applied in this thesis, the distinction proposed by Gough *et al.* [50] is more meaningful. Gough *et al.* [50] distinguish between aggregative and configurative synthesis methods, as depicted in Figure 1.2.

Aggregative reviews aim to test a theory or hypothesis and to identify the magnitude and variance of an effect. It therefore collects homogeneous forms of evidence, aiming for high comparability of reviewed studies. Such evidence is often collected from studies with homogeneous research designs. The synthesis then aims to *aggregate* the evidence by averaging the findings from the primary studies and assessing their variation and consistency [50].

A configurative review, on the other hand, aims to develop concepts or theory by interpreting and exploring patterns in evidence. Configurative synthesis methods are capable of a larger heterogeneity in the forms of the reviewed evidence. It triangulates different lines of evidence to develop and refine theory. The broader scope of eligible evidence does commonly not allow for an exhaustive search for evidence, as would usually be the standard for aggregative reviews [50].

Systematic reviews usually include elements of both synthesis approaches. This also applies for the two synthesis methods I use in this thesis. Still, if aggregative and configurative synthesis approaches would form a one dimensional scale, the meta-analysis, I conduct in Chapter 3, would be placed far on the aggregative side of the scale, while the realist synthesis, in Chapter 4, would be placed much more towards the configurative side. For the meta-analysis, I collect homogenous estimates of effect sizes, which can be harmonised and averaged across studies. For the realist synthesis, I collect, test, and refine hypothesised theories based on various lines of evidence that cannot easily be harmonised. The meta-analysis, however, also includes configurative elements, in particular, in the assessment of heterogeneity in the primary evidence.

For the meta-analysis, I collect quantitative estimates of emissions reductions caused by carbon pricing policies. I synthesise these estimates to study the emissions reductions achieved by carbon pricing schemes implemented in different jurisdictions.

For the realist synthesis, I collect hypotheses and evidence on how and under what conditions carbon pricing triggers emissions reductions. I synthesise the evidence to derive an evidence-based theory of the mechanisms and context factors determining the effectiveness of the policy.

1.4 Research questions

This thesis is guided by three overarching research questions on the feasibility and usefulness of evidence synthesis for evaluating climate policies.

- What are the benefits of applying evidence synthesis methods to climate policy evaluations?
- What are the remaining challenges in mainstreaming evidence synthesis in climate policy evaluation?
- How can evidence syntheses be used to inform actors in the different climate policy development spheres, e.g. global environmental assessments, researchers and policymakers?

In addition to these overarching questions, each of the syntheses conducted in Chapters 2 to 4 sets out their own research questions. The overarching questions are addressed and discussed in Chapter 5.

Chapter 2

Living evidence and gap map of carbon pricing policy evaluations

2.1 Introduction

Comprehensively tracking the available scientific evidence and making it easily accessible is a critical starting point for learning about climate policies in science and policy. Climate policy research is conducted across disciplines and countries, often with limited overlap between research fields. Publications are scattered across hundreds of disciplinary journals, do not follow any common standards in the use of keywords, and often differ in their jargon, making it particularly difficult to trace via the available searching tools. The evidence synthesis community has developed evidence gap mapping as an important scientific practice to systematically gather and lay-out scientific evidence in a particular field [86, 87]. A comprehensive, continuously updated database of all research relevant to a research area, pre-sorted by the fundamental study characteristics, would simplify the interdisciplinary assessment of the state of research.

The best available map of climate policy evidence is provided by Callaghan *et al.* [15], mapping out the entire evidence on climate mitigation policies from 84,990 research articles into 48 categories of policy instruments. The map further distinguishes between ex-post and ex-ante research methods and by the studied sector. This is a valuable overview of which types of climate policies have been studied, but does not provide any insights, what it is about a policy that is studied in each article. I here draw upon the more fine-grained method of evidence and gap mapping [82, 86, 87], aiming to provide a more detailed overview of policy evaluations on carbon pricing, studying which policy outcomes are studied, for which schemes, using what research methods. This allows for a rigorous assessment of what research questions have yet been addressed and where research gaps remain in the available evidence.

In this chapter, I provide the first evidence and gap map in the field of climate policy evaluation with an application to the evidence on carbon pricing policies. Together with a group of co-authors, we systematically search three bibliographic databases for empirical ex-post literature on carbon pricing, screen more than 50,000 scientific publications using a machine-learning assisted screening approach. We categorise the identified 4,054 relevant publications based on the geographical location of the implemented policy, policy outcome, study method, and emission sector. We use the substantial number of more than 10,000 hand-labelled documents to test how artificial intelligence can automatically identify and categorise relevant documents in future and evaluate its performance against manual classifications. We share the identified carbon pricing assessments in a comprehensive, living, open-access database where everyone can use and explore the relevant literature to support efforts to strengthen evidence-based policy as well as IPCC assessments through rigours evidence synthesis. We demonstrate the direct value for science, policy, research prioritisation and IPCC assessment by identifying research gaps in the primary evidence and evidence synthesis gaps.

The interactive evidence and gap map is available here: `https://climateliterature.org/#/project/carbonpricing`.

2.2 Methods

In this study, we apply an evidence and gap mapping approach [82, 86, 87] to provide a systematic overview of the empirical literature on carbon pricing, focussing on ex-post policy assessments. Following the guidance by the Campbell Collaboration [26, 130], we search a broad range of relevant literature, screen for eligibility, extract data, and synthesise the information. The research methods were set out in a research protocol [135]. In the remainder of this section, we summarise our search strategy, the screening and coding process, our synthesis method, and potential methods for making this map "living".

2.2.1 Literature search

We conducted a comprehensive search for potentially relevant literature in the bibliometric databases Web of Science, Scopus, and OpenAlex. The search string focuses on the studied intervention, without any restrictions for other dimensions of the PICO (population, intervention, comparator, outcome) structure, which is usually suggested for the search string development [130, 136]. The use of a broad search is facilitated by the application of machine-learning at the screening stage. The comprehensiveness of the search string was tested against a benchmark list of known relevant studies compiled from five carbon pricing reviews [18, 137–140]. The final search string, translated to the query syntax of each data source, is provided in the protocol [135]. The search was performed in June 2023.

The search results are imported into the NACSOS review platform for screening [141]. During the import, the platform automatically removes duplicates based on the similarity of title and abstract and other bibliometric meta-data. After deduplication 50,406 unique publications remain (see Figure 2.1).

2.2.2 Screening

We screen the 50,406 publications found by the literature search for relevance based on three criteria: (i) the type of policy, (ii) whether the policy is implemented, and (iii) the use of empirical ex-post study methods. Included studies should evaluate a carbon pricing policy in the form of carbon taxes or cap-and-trade schemes that apply a uniform price across emissions sources based on the carbon dioxide content of the emissions or, for other greenhouse gases, their equivalent. This does not include purely voluntary schemes or policies that introduce carbon prices through offsetting requirements. We only include studies that analyse observed data from policies that

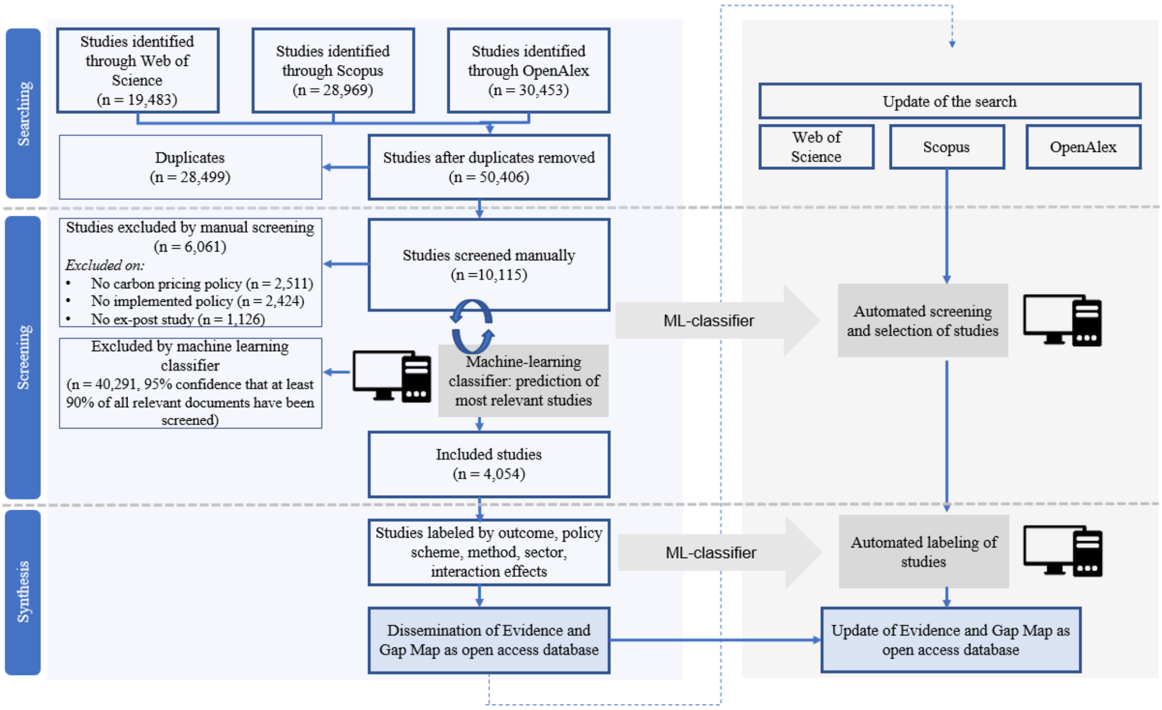

Figure 2.1: **Flow diagram of the literature search, screening, and automation**: The left panel is adapted from the ROSES flow diagram for systematic reviews [142] and depicts for each stage of the search and screening process, how many documents were seen and included. The right panel depicts how the living evidence map can be updated in future.

are already implemented with a focus on empirical policy assessments using ex-post study designs. We are inclusive of all sorts of qualitative and quantitative ex-post study methods but exclude purely theoretical or modelling studies.

We substantially speed up the screening and data extraction process by only considering the title and abstract (not the full text) and by prioritising which documents to screen using machine-learning. The focus on title and abstract information comes at the expense of detail in the characterisation of the available evidence, but allows coding not only hundreds, but thousands of studies. Furthermore, it facilitates automated updates of the evidence and gap map in the future as this information is readily available in bibliographic databases. The use of natural language processing to screen documents in their order of predicted relevance has shown to substantially reduce workloads without significant loss of recall or precision [91, 143, 144] and allows for the use of a less restrictive search query.

We screened an initial set of 1,276 documents of randomly selected articles, of which 105 were included and 1,171 excluded after resolving eventual disagreements between annotators. Based on this data, we train machine-learning classifiers to predict the relevance of unseen documents. We use these scores to iteratively select the next batch of documents for annotation that are most likely to be relevant. The cycle of training and predicting is repeated until all relevant documents are seen.

The first 3,272 abstracts were screened by at least two coders and discrepancies in inclusion decisions discussed with the research team. Once high agreement between the coders was reached (Cohen's kappa > 0.75), the screening was conducted by single coders and uncertainties were brought up in the research team.

The machine-learning classifier we use, is trained at regular intervals before assigning the next batch for annotation. At each step, we choose the best model and hyper-parameter setup by cross-validating performance of already annotated documents. In earlier rounds, simple Ridge classifiers, support vector machines, or logarithmic regression on dimensionality-reduced TF-IDF vectors performed best. The latter third of assignments was done using a fine-tuned BERT transformer model. Interestingly, domain-specific models like ClimateBERT [145] or SPECTER [146] did not seem to perform better than general and smaller models.

We used a robust statistical stopping criterion to decide when we could stop screening. After we had manually screened 10,115 documents, the conservative stopping criterion provided by Callaghan and Müller-Hansen [144] informs us that we have found at least 80% of the relevant documents at a confidence level of 80%. The same authors provide in an R package [147] a more realistic stopping criterion based on a non-central hypergeometric distribution [148], which reflects that the documents have not been screened at random. Applying the non-central hypergeometric distribution with a bias parameter equal to six, we are confident that we have found at least 90% of the relevant documents at a confidence level of 95%. We are thus confident that our evidence and gap map includes most of the relevant policy assessments of carbon pricing.

In total, we identified 4,054 ex-post carbon pricing policy assessments via the screening (see Figure 2.1).

2.2.3 Data extraction

To provide a structured overview of the identified carbon pricing studies, we developed a framework for the evidence and gap map, with the two primary dimensions being the specific policy scheme and policy outcomes studied. The World Bank [29, 149] lists 73 carbon pricing schemes implemented by different jurisdictions worldwide, which we use as our benchmark list of policies. The World Bank data furthermore provides us with additional information on the geographic location, implementation date, price level, and emission coverage of each of the policies that we use for our assessment.

As a second dimension we capture the policy outcome studied in each publication. We searched the literature for a suitable typology of climate policy outcomes, but found no typology that is at the same time comprehensive and fine-grained enough to capture all relevant policy outcomes. The IPCC provides a broad typology capturing six outcome dimensions: Environmental effectiveness, Economic effectiveness, Distributional effects, Co-benefits and negative side-effects, Institutional requirements, and Transformative potential [9]. For our purpose this classification of outcome domains was too broad to provide a detailed overview of the policy outcomes studied in the literature. We therefore developed a dedicated typology of policy outcomes for this evidence and gap map, synthesising the IPCC typology with other outcome classifications proposed in the literature [138, 140, 150–153] (see Appendix Figure A.1). The resulting typology of policy outcomes aims to balance between the particular requirements for the map on carbon pricing policies and the generalisability for evidence syntheses on other climate mitigation policies. The resulting typology captures the 14 outcome categories presented in Table 2.1.

We use three additional dimensions, referred to in the evidence and gap mapping terminology as *filters*. We use filters for the study method, the studied sector, and interactions with other policies. We found that abstracts usually allow us to identify if studies use one of the following methods: quantitative assessments, interviews or surveys, or literature review methods. Other study methods, in particular qualitative research methods, can often not clearly be identified from the abstract. For quantitative methods we further distinguish if statistical methods are used or if

Policy outcome	Description
Environmental effectiveness	Captures reductions (or other developments) in greenhouse gas emissions or energy use. Changes in energy use can be in the form of changes in the quantity of energy used or in the form of changes in the sources (coal, gas, renewables, etc.) of energy used.
Leakage	Captures the effect of the policy on relocations of emissions and production processes to other geographies or actors not covered by the carbon price.
Innovation & Investment	Captures the effect of the policy on research and development (R&D), demonstration of or investments into new (green) technologies or processes. Articles in this category could study efforts or outcomes of innovation processes.
Firm behaviour & Economic structure	Captures the effect of the policy on the behaviour, stocks and capacities of firms or the economy as a whole (which are not captured by any other category). Articles in this category may study firm behaviour, supply of good and services, use of technologies, capacity of energy installations, etc.
Prices of goods and services	Captures the effect of the policy on prices (e.g. energy prices). All assessments of price developments should be captured here, except for developments in the carbon (allowance) price. (Changes in firm value are also not captured here, but in the category 'Competitiveness'.)
Household behaviour	Captures the effect of the policy on the behaviour of individuals (households, consumers, etc.). This does not include the behaviour of firms or governmental actors.
Competitiveness	Captures the effect of the policy on the competitiveness of firms or entire economies. On firm level this includes output and productivity measures as well as financial indicators, such as profits, costs, access to capital, financial performance etc. On the level of economies this includes output (GDP), productivity etc.
Employment & Labour market	Captures all effects of the policy on (un-)employment, wages etc.
Distributional effects & Fairness	Captures the effect of the policy on the distribution of income as well as other social outcomes such as access and affordability of goods, poverty, well-being, or conflict. This category captures all dimensions of social outcomes or fairness concerns with respect to individuals (not firms).
Cost effectiveness & Efficiency	Captures whether the policy efficiently delivers its aims, including evaluations of the market efficiency or market imperfections of the carbon market.
Implementation process & Feasibility	Captures evaluations of the implementation process and the related feasibility of the policy. This particularly includes evaluations of carbon price developments, carbon price expectations, compliance, distribution of allowances, use of off-sets, banking of allowances, administration of the policy as well as political economy considerations. (While these are, strictly speaking, not outcomes of the policy, they are of high relevance for the evaluation of the policy.)
(Public) Perception	Captures the perception of the general public or specific groups towards the policy.
Environmental and health co-benefits	Captures the effect of the policy on environmental or health outcomes other than the reduction in greenhouse gas emissions. For environmental outcomes, this includes, e.g., effects on air or water quality, biodiversity, soil conservation, wildlife habitat, etc. For health outcomes, this includes effects on physical and mental health – including e.g. on fertility, (child) mortality, prevalence/incidence of diseases, etc. (The impact on socio-economic dimensions related to health, e.g. access to healthcare, should only be captured under "Distributional effects & fairness".)
Other	Captures any other policy outcome that cannot be allocated to any of the above categories.

Table 2.1: **Typology of policy outcomes**: The table provides definitions of the policy outcome categories, used in the classification of policy evaluations. The typology results from a synthesis of available outcome typologies [9, 138, 140, 150–153].

quasi-experimental designs are applied. For review articles that could potentially provide a systematic review we read the full-text of the article and look for a clear description of a search strategy and synthesis method to determine whether a systematic review was conducted. A second filter captures the emission sector the study focusses on, if applicable. We use the sector definitions from IPCC working group three [2] and distinguish between Energy, Industry, Transport, Buildings, AFOLU (agriculture, forestry, and other land use), and International Aviation and Shipping.

The information on all mapping dimensions and filters is extracted from the titles and abstracts of each relevant article. In line with the screening strategy, we first coded the information jointly in teams of at least two coders, discussed and resolved discrepancies and arrived at a collective understanding of the definitions of each dimension. After the first 1,712 documents were labelled, we reached a good agreement between all coders and continued with single labelling. We assessed the agreement for the single choice methods label using Cohen's kappa (>0.85) and for the multi-choice dimensions outcome, policy scheme, and sector we computed the mean agreement relative to all labels assigned by at least one coder (>0.85).

2.2.4 Synthesis

We use the collected data to assess the state of the research on carbon pricing policies. We assess the patterns of published studies for different policy schemes and policy outcomes. We study these patterns with respect to the study methods and the time of publication. We provide an overview of the identified literature as an interactive online tool, allowing for an exploration of the data in more detail with all the labelled dimensions and filters.

We study where gaps in the literature are, where clusters of primary evidence exists and whether these are synthesised by systematic reviews. We compare the gaps and clusters of evidence with features of the studied policies. In particular, we use data provided by the World Bank [29] on the amount of GHG emissions covered by each scheme, their respective carbon price level, the duration since the policy was implemented as well as the design of the policy as a carbon tax or ETS to understand whether these factors are able to explain which policy schemes receive higher or lower scientific attention.

2.2.5 Machine-learning classifier for automated updates

To avoid that the compiled map gets rapidly outdated by the exponential increase in relevant literature published after this study was conducted, we evaluate the potential of machine-learning to automate future updates of the map. As the manual screening and coding of literature is a labour-intensive exercise, we test to what extent and at which quality automated classifiers can be utilised to identify relevant research in line with our inclusion criteria and to classify the studies' analysed policy scheme, outcome, research method, and sector. We evaluate the quality of the automated classifiers for each sub-task based on the F1-score, which captures the harmonic mean of precision and recall. More details are included in a simulation of this process below. The first step of the update pipeline would retrieve the latest publications that match our search query from the three bibliometric databases. These will then be filtered according to our inclusion criteria using a machine-learning classifier. All included documents are enriched with labels for the studies' method, studied outcomes, sector, and policy scheme.

For the inclusion criteria we use transformer-based machine-learning classifiers trained on our entire hand-labelled set of 10,115 documents. For the method, outcomes, and sectors, we use similar multi-class or multi-label classifiers that are trained on our labels of the 4,054 included documents. For some of these filters we found logistic regression models to work best, on others transformer-based models work best. Each classifier is 8-fold cross-validated to estimate the variance in performance and potential biases. The large number of policy schemes, for some of which we only have few labels, prevents machine-learning classifier to work accurately and reliably. To this end, we found a dictionary approach to work best. We found that the policy schemes can usually be identified based on mentions of their respective geographic location and additional hand-crafted keywords in the text. For a slight simplification of the task, we group the policy schemes in China, Mexico, Canada and the US on the country level. The list of keywords is provided in the Appendix Table A.1.

2.3 Results

2.3.1 A large and fast-growing evidence base on ex-post assessments of carbon pricing schemes

Our evidence map reveals a large amount of ex-post assessments of carbon pricing policies in the scientific literature. We identify 4,057 relevant studies for 62 of the 73 implemented carbon pricing policies. During the period of the IPCC's sixth assessment cycle (2014-2021) alone, more than 1,950 new ex-post carbon pricing studies have been published. Only 134 of these are cited in the IPCC assessment report.

The growth in the ex-post evidence on carbon pricing is exponential and even faster than the average growth of the climate literature. Since 2005 average annual growth of publications on carbon pricing was 17%, compared to the already large growth rate of 15% per year for the climate literature as a whole. This is not surprising as the ex-post literature on carbon pricing only started to take off during the second half of the 2000s when carbon pricing schemes started to be adopted more widely (see Figure 2.2). Growth since 2019 has been particularly fast, with an average annual growth of 39%. In 2022 the research interest in the topic grew by more than 75% compared to 2021. If this trend continues, it highlights the importance of tracking this dynamic literature base to keep abreast with the evidence for IPCC and other science assessments as well as evidence-based policy in general. More importantly, a rapidly expanding evidence base further emphasises the need for systematic assessments and syntheses of the evidence base.

Emission trading schemes (ETS) receive significantly larger research interest than carbon taxes. About 78% of the ex-post carbon pricing assessments study an ETS compared to only 5% focussing on carbon taxes. The remaining policy assessments study multiple carbon pricing policies across countries without focussing on a specific scheme (8%) or the policy cannot clearly be identified from the title and abstract (8%). The higher attention for ETSs can be partially explained by the larger emission coverage of these policies. But we find 15 times as much literature on ETSs

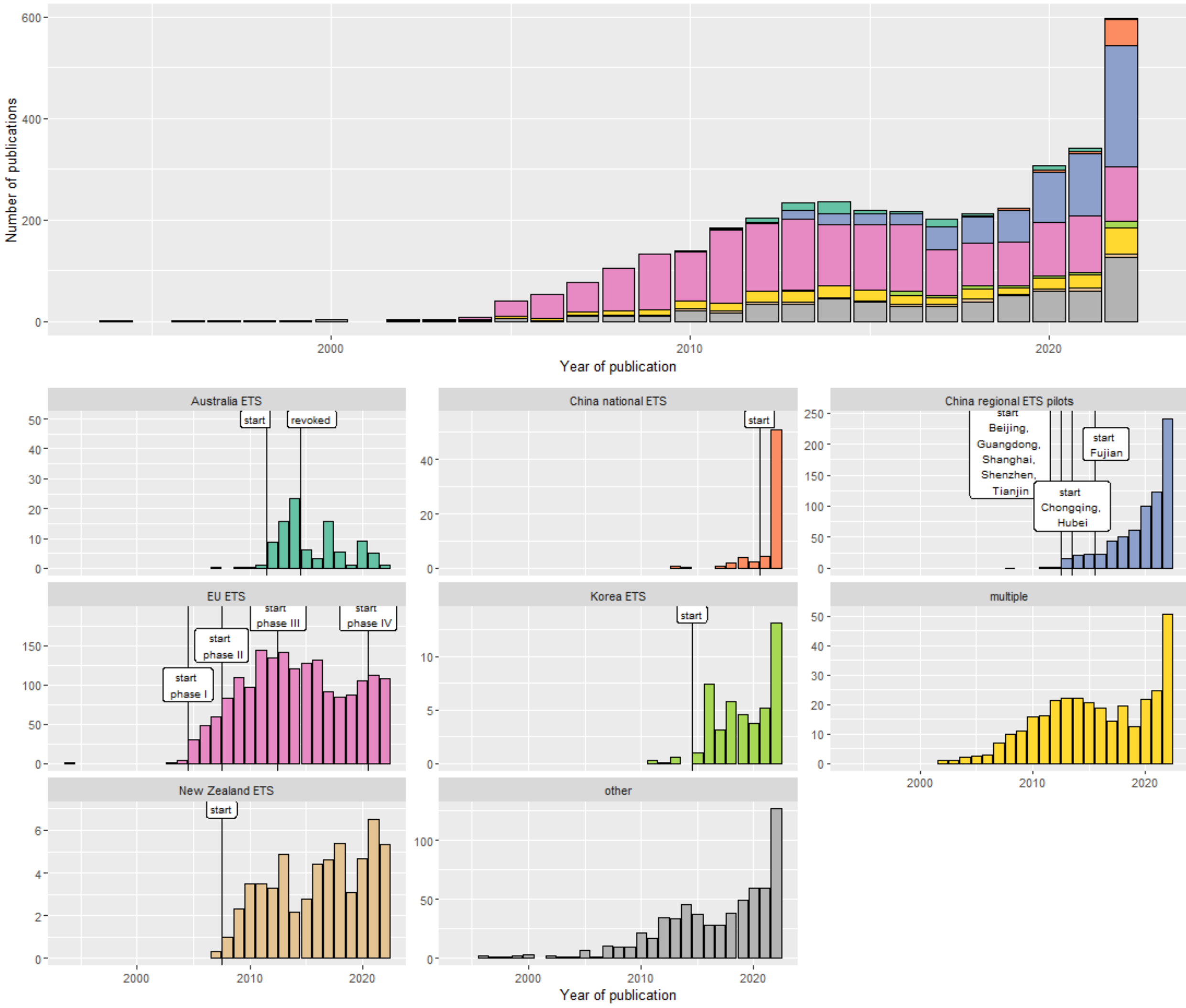

Figure 2.2: **Ex-post carbon pricing policy evaluations by year of publication**: The upper panel shows the development of all ex-post carbon pricing assessments across time. The lower panels depict the literature published by carbon pricing scheme and indicate relevant events for the respective schemes. Carbon pricing schemes with less than 50 ex-post policy evaluations are collected in the panel labelled "other".

than on carbon taxes, while their emission coverage is only three times as high as for carbon taxes [149].

Research interest in ex-post assessments of carbon pricing schemes is highly focussed on the three largest ETSs: the supra-national ETS in the European Union (EU ETS), the national ETS in China, and the eight sub-national ETS pilot schemes in China (see Figure 2.3). Together, they attract 65% of the total research activity in this field, while the remaining 70 schemes are only covered by the remaining 35% of the literature. Most ex-post policy assessments are conducted for the EU ETS (1,876), which was also the first ETS ever implemented. The ETS pilots in eight Chinese provinces are usually studied collectively and account for 845 policy assessments. Even though, the national ETS in China was only introduced in 2021 it already accounts for the third largest amount of published policy assessments (100). The EU ETS, Chinese national ETS, and the Chinese regional ETS pilots also account for the largest shares of GHG emissions covered by carbon pricing schemes, with 12%, 39%, and 7%, respectively. Also the ETSs in Korea, New Zealand, and the special case of Australia, where an ETS was implemented and revoked only 2 years later, receive substantial scientific attention, with more than 50 ex-post policy assessments each. From the 36 implemented carbon taxes, the ones in British Columbia (49) and Sweden (37), receive the highest scientific attention.

A large number of carbon pricing schemes have been rather neglected in the scientific literature. For 33 carbon pricing schemes, we find less than five ex-post policy evaluations, of which eleven are not covered by a single study — frequently referred to as an absolute evidence gap [86]. Even five studies on a given carbon pricing scheme can hardly be expected to answer all policy relevant research questions on the intended and unintended policy outcomes, process evaluations and equity concerns, not even considering that rigorous and nuanced policy learning usually requires the application of multiple study designs. Figure 2.4 shows no clear pattern, which schemes receive very low scientific attention. While the majority of the neglected schemes was introduced less than five years ago, the list also includes the carbon taxes in Poland, Slovenia, Estonia and Latvia, which have been implemented between 1990 and 2004. The understudied policies also include carbon pricing schemes

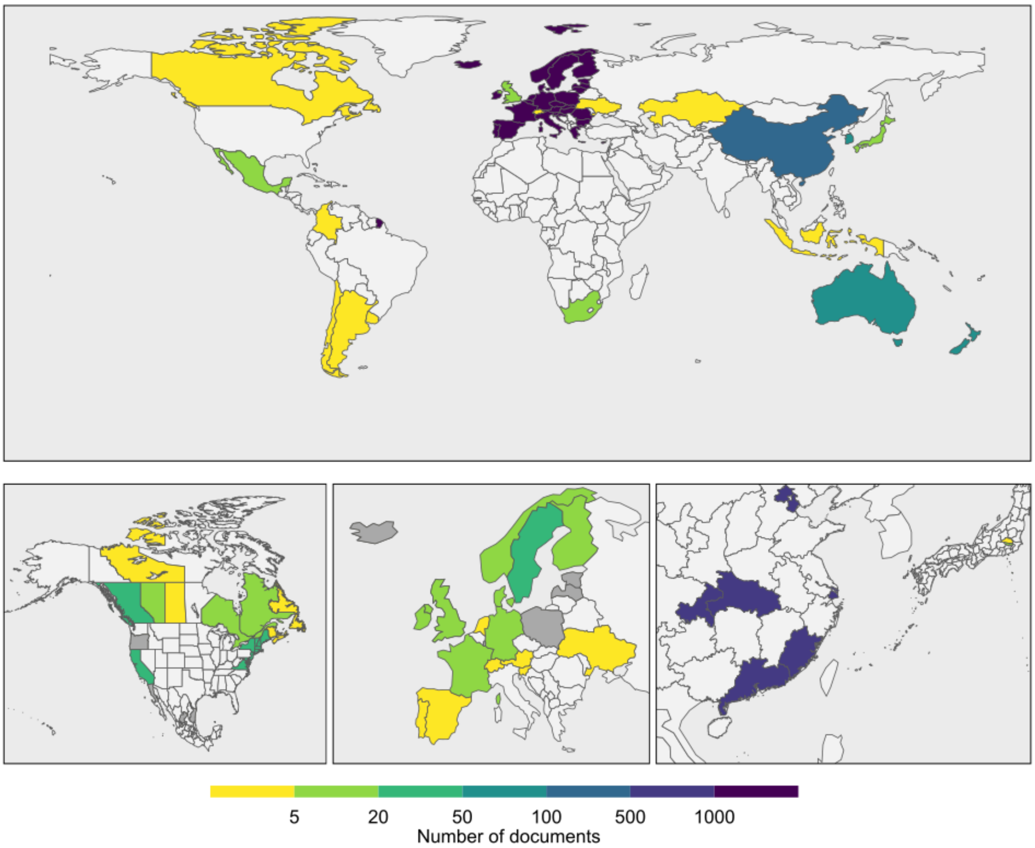

Figure 2.3: **Number of ex-post policy evaluations per implemented policy**: The upper panel depicts national carbon pricing schemes as well as the EU ETS. Other European carbon pricing schemes are not depicted in the upper panel. The lower panels depict (from left to right) regional carbon pricing policies in Canada, USA, and Mexico; national carbon pricing policies in Europe; and regional carbon pricing policies in China and Japan. Countries or regions coloured in grey have a carbon pricing policy implemented, but no ex-post policy evaluation was identified.

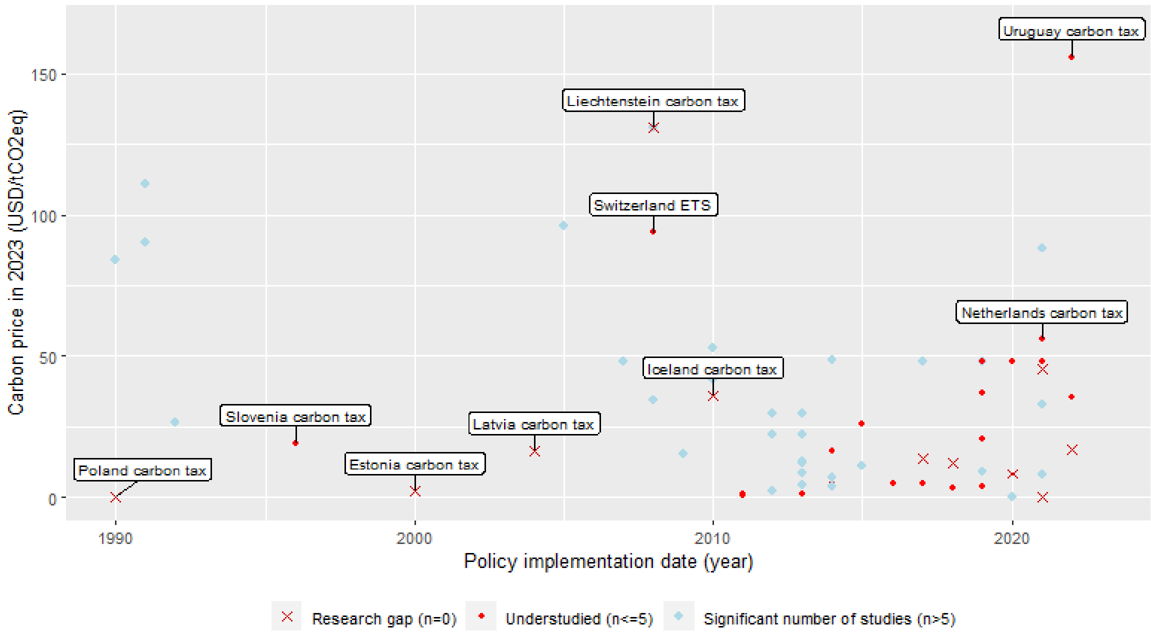

Figure 2.4: **Research gaps and understudied carbon pricing schemes follow no clear pattern**: The carbon price level (y axis) or the time since the introduction of the policy (x axis) cannot explain that some carbon pricing schemes receive little or no scientific attention. We use data on carbon prices and implementation dates by the World Bank [29]

with considerable carbon prices (in 2023), like the carbon taxes in Uruguay (US$ 156), Liechtenstein (US$ 131), and Netherlands (US$ 56).

2.3.2 Shared attention to policy outcomes and implementation issues

In order to understand the available evidence base, it is critical not only to know which carbon pricing schemes have been studied, but also what aspects of these policies. Almost half of the available ex-post studies analyse the institutional requirements of the policy, i.e. the design, implementation and functioning of the policy in its institutional context. The large research interest likely results from the novelty of cap-and-trade schemes, where the policy experience beyond carbon emission trading schemes is rather scarce. The research in this area includes assessments of the allocation of allowances, use of off-sets, participation in the secondary market, trading volumes, formation of carbon prices, and other market features of ETSs. In addition, about 1.4% of the policy assessments study the perception of the society towards carbon pricing.

The other half of the literature is made of policy outcome evaluations. Policy outcomes describing the transformative potential of the policy receive the highest research interest (19%), followed by economic (14%) and environmental (13%) outcomes (see Figure 2.5). Significantly less published research focusses on the distributional effects (1.1%) of the policy as well as on co-benefits and negative side-effects (0.7%).

Studies on the transformative potential assess how firms and individuals react to the carbon pricing policy. This research has focused on how firms adopt lower carbon technologies and practices (firm behaviour and economic structure, 7.7%), invest in innovations (5.3%), or adapt their prices for carbon intensive goods (1.4%). Fewer studies (0.5%) investigate changes in household behaviours.

Environmental benefits and economic costs receive an about equal scientific attention. Reductions in greenhouse gas emissions or energy use are studied by 11% of the ex-post policy evaluations and 1.6% assess to what extend emissions are relocated to territories or firms not subject to the carbon pricing policy. Whether these emissions reductions are achieved cost effectively is studied by 3.9% of the studies. Significant evidence is available to answer the question whether the climate policy has a positive or negative effect on firms (10%) and the labour market (0.5%).

A comprehensive assessment of complex climate policy questions requires the use of a wide range of qualitative and quantitative research methods. In ex-post carbon pricing assessments about 40% of the studies use statistical inference methods, of which a quarter applies quasi-experimental methods such as difference-in-differences or regression discontinuity designs. 3.8% conduct interviews or surveys and 3.3% are reviews of the literature. The remaining 53% of the studies use qualitative or descriptive research methods. While Figure 2.5 shows that for some policy outcomes the evidence is dominated by quantitative research findings, the broad picture reveals that policy questions on each of the outcome dimensions are answered by a mixture of quantitative and qualitative policy evaluations.

The research methods used to study carbon pricing policies differ across schemes. Among schemes, with more than 30 identified studies, both the Chinese national ETS and the regional pilot schemes stand out, with respectively 51% and 68% of

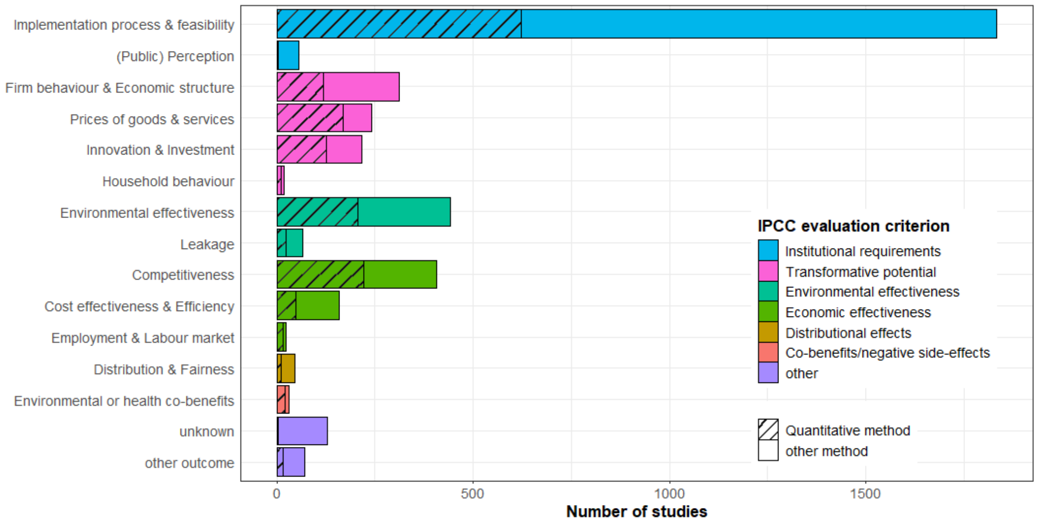

Figure 2.5: **Number of ex-post policy evaluations by studied policy outcome**: Each bar depicts the number of policy evaluations identified for each policy outcome. Studies assessing more than one policy outcome are weighted by the number of policy outcomes assessed. The bars are ordered and coloured reflecting the IPCC typology for evaluation criteria. Studies using quantitative research methods (quasi-experiments or statistical inference) are presented by the crosshatched areas.

studies using statistical inference methods. For the Korean ETS the number of studies using statistical inference are close to the average of 40%. For all other carbon pricing schemes the share of statistical assessments account for less than 40% of all ex-post evidence. For cases, such as the ETSs in Australia and California this goes as low as 8 and 9%, respectively. The Chinese regional ETS pilots do not only stand out for providing the largest share of quantitative assessments, they also provide 302 quasi-experimental studies, while for all other schemes combined only 96 quasi-experimental studies are conducted.

While 3.3% of all studies are reviews of the literature (n=134), we identify only a single systematic review of ex-post evidence on carbon pricing policies. It synthesises the evidence on financial risks in the markets for carbon allowances of ETSs [154]. The low number of systematic reviews indicates a large evidence synthesis gap on carbon pricing policies.

2.3.3 Critical evidence and evidence synthesis gaps in carbon pricing research

Figure 2.6 provides an overview of the evidence on carbon pricing, i.e. how many ex-post assessments are available for individual carbon pricing schemes and which outcomes are studied more or less frequently. First to note is the very uneven distribution of carbon pricing studies both across schemes and outcomes (see sections 2.3.1 and 2.3.2 above). Studies cluster across both dimensions pointing towards inefficiencies in the evidence ecosystem and potential wasteful research practices, where many scholars tend to study similar subjects whilst leaving many evidence gaps.

There are many mature areas of primary ex-post evaluations of carbon pricing schemes featured by dozens or even hundreds of studies of a single outcome by a single carbon pricing scheme. Because our mapping exercise also highlights the gulping absence of rigorous evidence syntheses on carbon pricing with only one systematic review, these mature areas of research mark evidence synthesis gaps, where the wealth of primary research on carbon pricing has not been rigorously synthesised for evidence-based decision-making. Systematic reviews would particularly be needed to assess the available carbon pricing evidence on environmental effectiveness, innovation and investment, competitiveness, impacts on prices of goods and services, or firm behaviour. Syntheses of the vast amount of evidence on the implementation process could provide robust answers to policymakers on some fundamental questions arising in designing an effective and equitable carbon pricing scheme. Comprehensive efforts to summarise this heterogeneous body of qualitative and quantitative evidence would require using the whole portfolio of evidence synthesis methods [27] – even though most qualitative and mixed evidence synthesis methods have been hardly applied at all by the climate policy evaluation community.

But there are also significant and pressing primary research gaps, calling for intensified research efforts in specific research areas. As a considerable number of carbon pricing schemes have not or only barely been evaluated, there are outcome dimensions that have been neglected – even for some of the more well researched carbon pricing schemes. Many basic research questions remain unanswered and mark absolute evidence gaps: How effective is the Argentinian carbon tax in reducing emissions? What impact did the carbon taxes in British Columbia, Denmark, Finland, or Japan have on investments and innovation? Or how did the taxes in Sweden or Norway influence the prices of energy and other goods? Looking at Figure 2.6

highlights that the absence of ex-post evaluation evidence is not an exception, but instead is rather common. Too little attention has in particular been given to questions on the distributional implications of carbon prices and potential co-benefits for human health or the environment. Across countries and carbon pricing schemes these are only studied by a very few empirical assessments, leaving policymakers without sufficient evidence on these socially important aspects of the policy.

2.4 Towards a living evidence base

Publicly available evidence and gap maps are critical to stay abreast with the available evidence in particular fields of research. They can help governments, NGOs, research funders, and researchers to easily identify relevant evidence, or gaps in evidence and evidence synthesis. This is useful across a variety of tasks including priority setting, horizon scanning or the generation of rigorous, synthetic evidence.

Traditionally, evidence and gap maps are snapshots in time and quickly outdated. For example, 15% of all studies in this evidence map on ex-post assessments of carbon pricing were published within the last year of the analysis (2022). We expect the available evidence to double in size in five years or less. This leads to the problem that the most recent evidence is not available when it is actually needed. The evidence synthesis community has reacted to this problem with the idea of "living evidence" [94, 96, 99, 155, 156].

Living evidence tackles the constant growth in published research by frequently updating evidence syntheses. Each round of updates searches for newly published literature and integrates it into the synthesis. This allows to keep up with new developments in the field instead of capturing the state of research at a single point in time [99].

To provide a data source for IPCC and other science assessments, as well as for evidence-based policy researchers, we aim to regularly update our database of carbon pricing evaluations. Given the size and growth of the literature, future updates by hand would require substantial human resources. In fact, during the COVID-19 pandemic – the mainstreaming of living evidence in health – this was achieved in a

Figure 2.6: **Evidence and gap map**: Each cell presents how many ex-post policy assessments study the outcome dimension depicted by the column for the carbon pricing scheme depicted by the row. The large number of empty cells reveals many absolute evidence gaps. The eleven carbon pricing schemes without a single policy evaluation (Baja California carbon tax, Estonia carbon tax, Iceland carbon tax, Latvia carbon tax, Liechtenstein carbon tax, Luxembourg carbon tax, Massachusetts ETS, Oregon ETS, Poland carbon tax, Tamaulipas carbon tax, Zacatecas carbon tax) are not depicted.

Filter	**Precision**	**Recall**	**F1**
Inclusion (yes/no)	0.76 (± 0.01)	0.76 (± 0.01)	0.76 (± 0.01)
Methods (6 classes, single choice)	0.76 (± 0.02)	0.68 (± 0.03)	0.69 (± 0.02)
Outcomes (15 classes, multiple-choice)	0.58 (± 0.01)	0.77 (± 0.01)	0.64 (± 0.01)
Sector (6 classes, multiple-choice)	0.83 (± 0.03)	0.79 (± 0.04)	0.80 (± 0.03)
Policy region (40 classes, multiple-choice)	0.70 (± 0.35)	0.74 (± 0.39)	0.71 (± 0.36)

Table 2.2: **Performance of machine-learning classifiers**: The performance of the classifiers for Inclusion, Outcomes, and Sector are measured as the average across an 8-fold cross-validation. The performance of the classifier for Policy region is a weighted mean across the 40 classes. Standard deviations are reported in parentheses.

huge, collective community effort. In the meantime, a growing discourse has emerged on how such periodic updates can be facilitated by utilising machine-learning. In the following, we demonstrate how well machine-learning classifiers could automatically update this evidence map.

To ensure the usefulness of our database, the automation must be rigorously validated and closely monitored to maintain a high level of quality. To do so we cross-validate the performance of machine-learning classifiers on our hand-coded data. Our experiments show satisfactory performance in selecting relevant studies based on our inclusion criteria with an F1 score of 0.76. We can correctly predict the relevance of a given study for our evidence and gap map in most of the cases. We list scores for all metrics and filters aggregated across results from our 8-fold cross-validation and respective classes in Table 2.2. Experiments with more complex transformer-based models suggest, that these can result in slight improvements (one to four percentage points F1) but require further investigation to ensure optimal performance for their application in an automated living evidence and gap map update. We provide reflections on what these performance metrics might mean for real-world applications in the discussion below.

We also simulate how classifiers would hold up over time without additional hand-labelling of incoming data. To do so, we assume this study would have been conducted in 2019 and only train the models on articles published up to this point. Since we have complete hand-labelled data for the following years, we can simulate automatic updates and measure the performance, respectively. In this scenario, the quality of the inclusion criteria remains stable throughout 2020 (F1=0.72), 2021 (F1=0.64), and 2022 (F1=0.71). The generally slightly lower performance in this simulation can be explained by smaller sample of studies the machine-learning is trained on. For the actual living map, we would still need to hand-label a sample of documents in regular intervals to maintain the reliability of our predictions.

2.5 Discussion

In this first systematic assessment of ex-post policy evaluations on implemented carbon pricing policies, we screen an extensive amount of scientific studies assessing the various policy outcomes of carbon pricing. More than 4,000 research articles have been published across the last three decades, evaluating the policy experience gathered for one of the most prominent climate mitigation policies that governs more than 20% of today's greenhouse gas emissions [149]. This evidence base is a valuable source to learn from past experiences for the choice and design of climate policy pathways leading into a carbon neutral future [5, 157].

We find a large and fast growing evidence base scattered across different research communities. In fact, literature growth is even faster than for the climate change literature as a whole and is seen to accelerate in more recent years. The scale and rapid expansion of the available literature has important implications: First, it is increasingly impossible for individuals and even groups of scholars to track this expanding evidence base. Second, we can learn from the existing ex-post evaluation evidence on what climate policies work, under what conditions, and why, by rigorously synthesising this evidence.

Our study reveals a huge evidence synthesis gap. The 4,000 primary studies are in strong contrast to only a single systematic review we identified via a comprehensive search and screening of the literature. Research on climate policies is almost exclusively summarised in traditional literature reviews, which do not apply methodological rigour as commonly done in primary policy evaluation research. There is a large literature across different scientific fields that show the different biases traditional literature reviews are prone to [19, 41, 158, 159]. Accelerated learning in science and policy on what climate solutions work will depend on the application of rigorous scientific methods that avoid misleading conclusions by systematically minimising bias.

Similarly, the prospects for enhancing policy evaluation in future IPCC assessments directly depend on efforts within the research community to rigorously synthesise this evidence base. For example, during the last assessment cycle (2014-2021) almost 2,000 relevant ex-post carbon pricing assessments have been published. This vast amount of research cannot be assessed by IPCC authors – let alone in a volunteer process – unless it is synthesised by the research community. We argue that building a rigorous evidence bank is critical for evidence-based climate policy [108] and it is a good moment in time to start synthesising key lines of evidence for the commencing seventh assessment cycle of the IPCC.

Our evidence and gap map identifies policy-outcome combinations in the field of carbon pricing with a rich underlying evidence base. These could be a starting point for a dedicated community effort to synthesise the evidence on implemented carbon pricing policies. In Chapters 3 and 4, I start filling some of the evidence synthesis gaps, conducting two systematic reviews on the environmental effectiveness of the policy. As outlined in section 1.2.3, the number of systematic reviews is slowly increasing, with a few reviews published since the cut-off date for this map. These efforts need to be significantly increased, considering the large amount of available evidence identified here.

But we also still find many gaps in the primary evidence. In fact, the evidence base is very clustered. Topics that are generally not well covered are the social outcomes of carbon pricing policies, including distributional and labour market effects, as well as potential co-benefits for human health or the environment aside from greenhouse gas emissions. This underscores the findings by Ohlendorf *et al.* [117] and Cuevas *et al.* [121], who identified substantial biases towards ex-ante studies for these research areas. The research also focuses largely on a few well studied carbon pricing schemes, while leaving many knowledge gaps for a large number of understudied schemes. This transparent overview of the available evidence should inform future research priorities. Filling one of the absolute evidence gaps might bring more novel insights than yet another study on the EU ETS or the ETS pilots in China. This does not mean that these policies are exhaustively studied, but points towards research topics that have yet received too little attention. Our map allows researchers and research funders to evaluate what evidence is already available to avoid wasting resources for an already answered research question [160].

There are many thousands of ex-post evaluations on other climate policy instruments. Our focus here is on the ex-post evaluation literature on carbon pricing policies. Callaghan *et al.* [15] suggest that there are overall more than 80,000 climate policy assessments (>40,000 of which are ex-post assessments), highlighting the need for similar efforts to map out and synthesise the evidence across policy instruments. In this study we adapted the method of evidence and gap mapping [26, 86, 130] to the field of climate policy evaluations. It proves to be a suitable tool to organise the empirical literature and allows to uncover previously unknown evidence gaps. The methodology can be applied to map the research on other climate policies. The typologies developed here can be used for the mapping of literatures on other policy instruments and the large number of hand-screened documents can serve as a pre-training dataset to better identify ex-post evaluation studies and reduce the required screening efforts.

The evidence and gap map compiled for this study is comprehensive, but it is still not complete. Some of the gaps may partially be explained by the focus of our study on research articles with an abstract published in English language. Integrating Spanish language literature, for instance, could potentially fill some gaps with respect to carbon pricing schemes in Central and South America. Literature maps commonly focus on a single language, but the extension to multiple languages could bring further benefits for comprehensive climate policy assessments. Our map also largely focuses on peer-reviewed scientific studies by searching bibliographic databases with limited coverage of grey literature sources. Policy evaluations by government bodies and international organisations are not exhaustively covered by our evidence map. These are mostly not readily accessible right now. It is a critical future task to start compiling such evaluations centrally. These tasks of further broadening the evidence map will be critical to better understand potential biases in the evidence base and to move towards a more integrated discourse on what climate solutions work, under what conditions and why across science and policy communities.

Our study highlights the potential of machine-learning towards living evidence maps. The rapid growth in climate policy assessments challenges the usefulness of any one-off evidence synthesis. We tested to what extent machine-learning applications can assist or replace human screening and labelling of the literature. When balancing between precision and recall, our algorithm is able to identify 76% of all relevant studies and every fourth included study will be a false positive, highlighting that artificial intelligence is not yet able to fully compete with human coders. Compared to the status quo, where it is practically impossible to keep up with the large amounts of relevant literature published every year, this still marks a step-change. With minimal human inputs, our evidence and gap map can be regularly updated, integrating the large majority of newly published carbon pricing assessments. Depending on the use of the data, the balance between precision and recall can also be adjusted. Researchers starting a systematic review may be more interested to retrieve all relevant documents and care less about falsely included studies that they can manually exclude. If one, for instance, wants to achieve a recall of 90%, the precision would go down to 71%. We contain classification scores in our automatically updated database, so that users can set a threshold that is appropriate for their use-case.

The quality of the automated identification of research articles can further increase with time. Users of this shared resource, may report any inconsistencies which will help to refine the original data-source and grow the number of high-quality hand-labelled data. The growing attention to large-language-models has also lead to experiments using prompts for screening and coding abstracts of research articles, however they have so far not reached the same performance as conventional machine-learning classifiers [161, 162]. More state-of-the-art transformer-based models have also shown slight quality improvements. However, we have not yet cross-validated all results across the vast number of general and domain-specific pre-trained transformers (e.g. SciBERT [163], ClimateBERT [145], SPECTER [146]). Over time, classifiers will need to be re-evaluated using hand-labelled data on the latest research in regular intervals to ensure that the automation still operates at the expected quality. Shifts in the language as research fields evolve may degrade the performance of trained classifiers. Furthermore, domain experts should review the search query to ensure that it still covers all relevant keywords.

We provide the (living) evidence and gap map as a publicly available, interactive database, listing all relevant research classified by the studied policy scheme, outcome, research method, and sector for an transdisciplinary audience. Evidence mapping is a valuable first step to foster evidence syntheses across climate policy instruments and outcomes. It can lead the way for the IPCC's seventh assessment cycle to build on the thousands of relevant studies that provide a detailed understanding of what climate policies work, under what conditions, and why.

Chapter 3

Meta-analysis on the effectiveness of carbon pricing

3.1 Introduction

Despite the experience from more than 70 implementations of carbon pricing schemes around the world [29] and the large amounts of ex-post policy evaluations discovered in Chapter 2, there remains no consensus in science nor policy as to how effective such policies are in reducing greenhouse gas emissions. Proponents have suggested carbon pricing as a key instrument to incentivise greenhouse gas emissions reductions on the basis that it would avoid the need for detailed regulatory decisions targeted at specific emission sources [164–168]. However, the effectiveness of carbon pricing is highly dependent on the context and the effect could be higher or lower based on the institutions and infrastructures [169, 170]. Critics doubt the ability of carbon pricing to unlock the investments required for the development and application of low carbon technologies [171]. There are also concerns about whether policymakers can overcome political barriers and raise carbon prices high enough to deliver emission reductions at the scale and pace required [171–173].

We aim to systematically review the empirical literature on the effectiveness of carbon pricing policies in reducing greenhouse gas emissions. While there are other market based policy instruments, such as fuel taxes, import taxes or value added taxes, we focus here on policies which impose a carbon price across fuels based on their carbon contents. One way to assess the effects of carbon pricing is to evaluate experiences in the real world. A growing scientific literature has provided quantitative evaluations of the effects of different carbon pricing schemes on emissions [18, 174, 175]. This evidence is usually provided in the form of quasi-experimental studies which assess the effect of the introduction of the policy (treatment effect). Based on this evidence, our meta-analysis addresses the question: What was the emission reduction effect of the introduction of a carbon price during the early years of its application? This is different from the question, how emissions respond to gradual changes in existing carbon prices. There exist only very few studies estimating this relationship between the carbon price level and emissions [176–178]. The comprehensive literature on the elasticity of fuel use in response to fuel price changes has been reviewed before in a number of meta-analyses [179–183].

We focus on the growing evidence base on the effectiveness of introducing a carbon price. Previous reviews of this literature have tended not to employ rigorous systematic review methods such as meta-analysis. A number of reviews describe the literature and summarise the findings of the primary studies but do not attempt a quantitative synthesis of the findings [18, 138, 139, 184]. Green [137] provides a range of effect sizes reported in the reviewed literature without any formal methodology for their harmonisation and analysis, concluding that the policy has no or only a very small effect on emission reductions (0-2%). None of the available reviews provide a critical appraisal of the quality of the primary studies considered. Biases of such traditional literature reviews have been widely documented in the literature [19, 185]. The lack of comprehensive systematic review evidence on a multitude of policy questions hampers IPCC assessments to learn from implemented climate policies [4, 5, 186].

We fill this gap by conducting a systematic review and meta-analysis of the empirical ex-post literature on the effectiveness of carbon pricing, covering 21 enacted carbon tax and cap-and-trade policies around the globe following the guidelines by the Collaboration for Environmental Evidence [187]. We use a machine-learning enhanced approach to screen 16,748 studies from five different literature databases, identifying 80 relevant ex-post policy assessments. We extract and harmonise estimates of average emissions reductions from the introduction of a carbon price. We conduct a meta-analysis on 483 effect sizes on 21 different carbon pricing schemes and estimate emission reduction effects. We study the heterogeneity in the reported findings and conduct a critical appraisal as well as a publication bias assessment to analyse the impact of different study design choices on the results. Our methodology is transparent and reproducible, ensuring that our analysis is updatable in the future as new information and experiences with carbon pricing policies are gained around the world [99]. The data and code is publicly available: `https://github.com/doebbeling/carbon_pricing_effectiveness.git`.

We find consistent evidence that carbon pricing policies have caused emissions reductions. Statistically significant emissions reductions are found for 17 of the reviewed carbon pricing policies, with immediate and sustained reductions of between -5% to -21% (-4% to -15% when correcting for publication bias). Our heterogeneity analysis suggests that differences in estimates from the studies are driven by the policy design and context in which carbon pricing is implemented, while often discussed factors like cross-country differences in carbon prices, sectoral coverage, and the design of the policy as a tax or trading scheme do not capture the identified heterogeneity in effect sizes.

3.2 Methods

In this study, we conduct a systematic review and meta-analysis to synthesise the evidence on the effectiveness of carbon pricing policies in reducing emissions. We broadly follow the guidance for systematic reviews by the Collaboration for Environmental Evidence [187], extended by a machine learning-assisted identification of relevant studies [91, 143]. We comprehensively search for relevant studies, screen for eligibility, extract information on estimated effect sizes, appraise and synthesise the data. A description of our methods has been published as a review protocol on OSF Registries in advance [188].

3.2.1 Literature search

The literature search closely resembles the search conducted in Chapter 2. If the evidence and gap mapping, conducted above, would have been completed at the time when we started this systematic review, the search strategy could have been guided by the evidence identified in the map for the *environmental effectiveness* category. Because of the chronological sequence of the two projects, the search for this systematic review was conducted independently of the evidence and gap map.

We search the bibliographic databases Web of Science, Scopus, JSTOR, RePEc and the web-based academic search engine Google Scholar using a broad search string which comprises a large set of carbon pricing synonyms and indicator words for quantitative ex-post study designs. The full query can be found in the protocol [188]. After the removal of duplicates the search, conducted in the second week of March 2022, returned a set of 16,748 articles (see Figure 3.1).

We screened these articles for their eligibility in two stages. First, we screened them at the title and abstract level using the NACSOS software [141, 189] followed by a screening at full text level. Studies are included if they infer a causal relationship between carbon pricing and the emissions development. Eligible studies analyse effects on emissions levels or emissions levels per capita. Studies were excluded if they assess the effect on emissions intensity or emissions productivity, i.e. the effect on emissions relative to output. The included policy measures are restricted to explicit carbon taxes and cap-and-trade schemes. Studies on implicit carbon taxes and carbon offsetting mechanisms are excluded. We only include studies published in English language.

The screening at the title and abstract level was simplified by an active learning algorithm, using support vector machines to rank the studies in the order of relevance. We stopped screening when we were 90% confident that we had identified at least 90% of the articles relevant to our systematic review, based on the conservative stopping criterion provided by Callaghan and Müller-Hansen [144]. This reduced the amount of manually screened documents by 77%. All articles included after the title and abstract screening were screened at full text, without any further application of

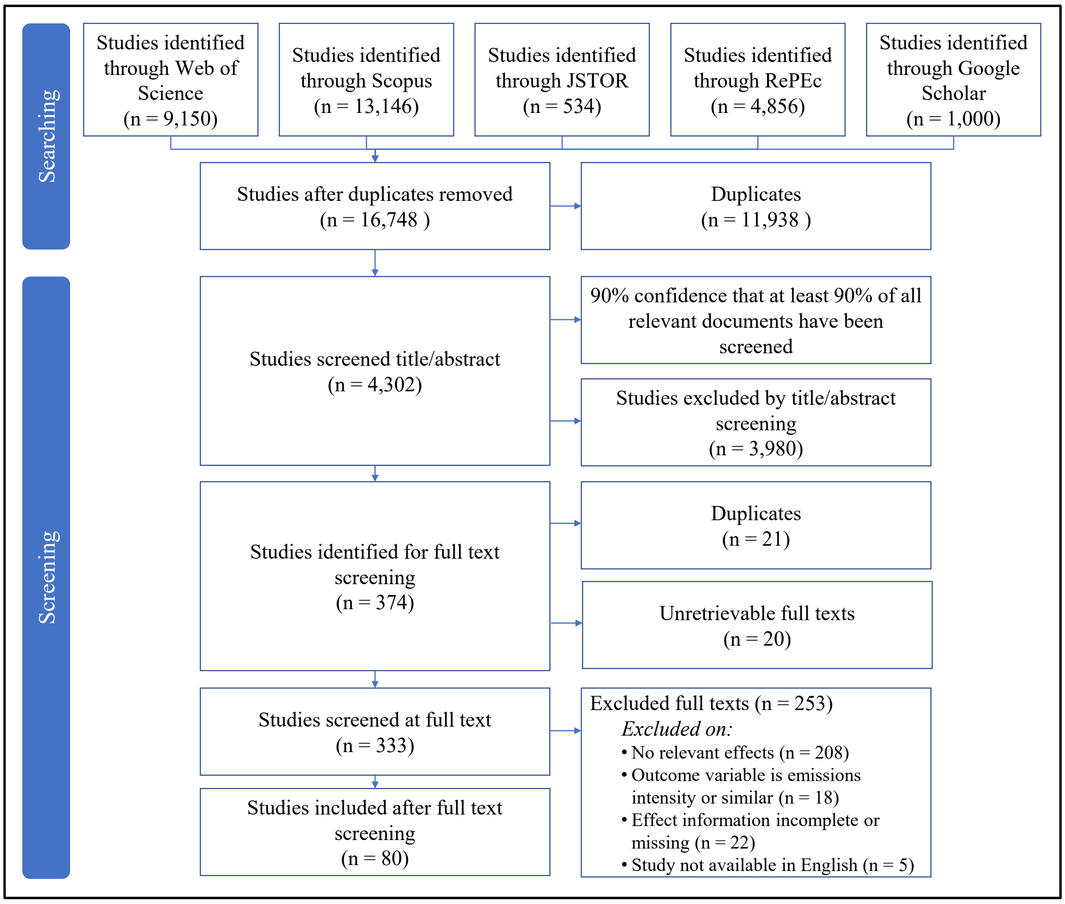

Figure 3.1: **Flow diagram of the literature search and screening process**: Adapted from the ROSES flow diagram for systematic reviews [190].

machine learning algorithms. Figure 3.1 depicts the articles included and excluded at each screening stage.

3.2.2 Data extraction and critical appraisal

From the included studies we extract the effect size information, including the estimated effect size and direction of the effect, the uncertainty measure, provided as standard error, t statistic, confidence interval, p-value, or the indicated significance level, as well as the provided mean emissions and, for price elasticity studies, the mean carbon price. We also capture information on the studied carbon pricing scheme, time of the intervention, study period, emission coverage (sectors, fuels, gases), study design, and estimation method.

We developed criteria for a critical appraisal, by adapting the ROBINS-I assessment criteria [132] to the specific nature of the research studies at hand. First, while the treatment (i.e. the policy application) in the reviewed studies is independent of the conducted research, the study design should cover a representative sample and suitable data. The control group needs to have high similarity with the treatment group, based on demographic, economic, and institutional proximity and similarity in pre-treatment emissions pathways. Statistical methods such as matching or synthetic control methods can increase the comparability of the control group with the treatment group. Second, the study design must control for confounding factors that are expected to influence the emissions of the study objects. For some studies we identify further risks of bias in the set-up of the statistical methods, which are also recorded.

All extracted data and the codebook used for the data collection are publicly available and can be accessed via: `https://github.com/doebbeling/carbon_pricing_effectiveness.git`.

3.2.3 Standardising effect sizes

We standardise the extracted effect sizes, based on the heterogeneous study designs and estimation methods, into a common metric. The largest part of the primary literature estimates treatment effects using quasi-experimental study designs (difference-in-differences or regression discontinuity in time). A few studies estimate the treatment effect by comparing the emission levels between countries with and without carbon pricing without any quasi-experimental design (termed *cross-country* studies in this review). Some studies estimate a carbon price elasticity, i.e. the effect of a marginal change in the carbon price on emissions. All effect sizes are transformed to treatment effects measured as a percentage difference between the counterfactual emissions without the policy and the observed emissions with the policy in place. Effect sizes expressed in tons of CO_2 are standardised using the mean emissions given in the study, while effect sizes from log-level regression specifications are standardised using exponential transformation. Effect sizes from price elasticity estimations are interpreted at the mean carbon price of the intervention during the period studied by the primary study.

Standard errors are derived accordingly. If the statistical (in)significance of an estimate at a specified significance level is the only uncertainty measure provided, this information is used to approximate the standard error. For the non-linear effect size transformation of log-level regression coefficients, we derive the standard error by keeping the t statistic constant. For effect sizes from price elasticity estimations, we interpret the standard errors at the mean price level, just as for the transformation of the effect size itself.

3.2.4 Effect size averaging

We use a multilevel random effects model to estimate the average treatment effect. The random effects model does not assume that all effect sizes converge to a common effect size mean [41], which in our case accounts for the heterogeneity in the studied countries and schemes. The common variance component is estimated using the restricted maximum likelihood (REML) estimation [191, 192]. We apply a multilevel estimation to account for the non-independence of effect sizes from the same study, assuming a compound symmetric variance-covariance matrix [192]. A robustness check using a fixed effects model specification is provided in the Appendix.

For the estimation of average treatment effects per carbon pricing scheme we extend the random effects model to a mixed effects model, inserting dummy variables for each scheme. Studies conducting a cross-sectional assessment of a set of carbon pricing schemes in multiple countries, are collected with a separate dummy variable. The eight Chinese pilot ETS schemes are collected in a single dummy variable, as they are commonly assessed together as a single policy in the primary studies. For many of the schemes only one to five studies are available, which does not allow for appropriate clustering of the effect sizes [193, 194]. The multilevel estimation of the model should still adequately capture the non-independence of effect sizes from the same study. Clustering of standard errors would have a marginal impact on the standard errors derived for the full sample averages (see Appendix Table B.2). The models are estimated in R using the metafor package [192].

To check that no single study exerts undue influence on the average effect sizes measured, we calculate Cook's distance and DFBETAS. For three studies in the sample the values of these metrics are distinctly different. All three studies assess the effect of emissions from the burning of coal. As these effects likely result from fuel switching without capturing the overall emission effect, 13 effect sizes from five studies with a focus on emissions from coal are excluded in the main assessment. An effect size average including these studies is provided in Appendix Table B.2.

We test for the presence of publication bias using a precision effect test [195, 196]. To correct for publication bias, we follow the guidance by [197] and [198] and estimate the model for a reduced set of the adequately powered effect sizes. To assess the power of each effect size we use the standard error of each effect and assume the genuine effect to be the average treatment effect from our full set random effects model. We follow common practise and assume studies with power of above 80% to be adequately powered [199]. We estimate a multilevel random effects model, in line with our main approach, instead of a fixed effects model proposed by [197] and [198].

3.2.5 Heterogeneity assessment

There is considerable heterogeneity in the effect sizes (I^2=0.86 in the random effects model). To capture the variation in the response to the policy, we code variables for the carbon pricing schemes as well as information on the sector coverage of the scheme (or the study, where the study focuses on a single sector), the mean carbon price level during the assessment period, and a variable distinguishing carbon taxes from cap-and-trade schemes. The information on sector coverage and the price level was added from external sources [200, 201]. We furthermore code a set of variables on the study design, estimation methods, and data used from the primary studies. The moderator variables are described in Appendix Table B.4.

Given the large number of potential explanatory variables, we use the Bayesian model averaging technique (BMA) [181, 202–204], employing a Markov chain Monte Carlo (specifically, the Metropolis-Hastings algorithm of the bms package for R by Zeugner and Feldkircher [205]) to walk through the most likely combinations of explanatory variables. In the baseline specification we employ the unit information prior which is recommended by Eicher *et al.* [206]. This agnostic prior reflects our lack of knowledge regarding the probability of individual parameter values. To test the robustness of our estimates we follow Havranek *et al.* [181, 204] and use the dilution prior that adjusts model probabilities by multiplying them by the determinant of the correlation matrix of the variables included in the model. Furthermore, as another robustness check, we follow Ley and Steel [207] and apply the beta-binomial random model prior, which gives the same weight to each model size, as well as Fernández *et al.* [208], who use the so-called BRIC g-prior. The BMA results using alternative priors are provided in the Appendix.

3.3 Results

3.3.1 Evidence from 21 carbon pricing schemes around the globe

With the help of our machine-learning assisted approach, we identify 80 quantitative ex-post evaluations across 21 carbon pricing schemes around the globe. Previous reviews covered a maximum of 35 research articles on the emissions reduction effect of carbon pricing policies [18, 137–139]. The systematic search conducted here proves to identify more than twice as many studies compared to any previous review. The full list of included studies can be found in Appendix Table B.1.

As shown in Table 3.1, the carbon pricing schemes covered here are very diverse and differ in terms of their specific policy design, scope, and policy context. For example, some of the schemes are targeted at large scale emitters in the industry and energy sectors, while others focus on households via home energy use and the transport sector. In the European Union, some sectors are regulated with a carbon tax while others are covered by the European wide emission trading scheme. We also observe substantial differences in carbon price levels of the covered schemes. All of these differences may give rise to considerable variations in emissions reductions achieved.

Beyond these differences in policy design, carbon price levels, and regional contexts, all considered policy experiences speak to the question whether carbon pricing is or is not effective in reducing greenhouse gas emissions. A systematic assessment and comparison of the outcomes of these policies can inform policymakers and future research by synthesising the available evidence.

The number of available ex-post evaluations on the effectiveness of carbon pricing differs substantially across schemes. Prior reviews suggested a bias towards evaluating schemes in Europe and North America [18, 137, 138]. However, here we find that the vast majority of the available ex-post evidence on the effectiveness of carbon pricing assess the pilot emission trading schemes in China – 35 of the 80 articles. There are 13 studies on the European emissions trading scheme (EU ETS), seven on the carbon tax in British Columbia and five on the Regional Greenhouse Gas Initiative (RGGI) in the United States. The remaining schemes are evaluated by a single or very few studies.

Our systematic review also reveals some fundamental evidence gaps in the literature. Despite the broad set of bibliographic databases searched, we found evidence only for 20 out of 73 carbon pricing policies in place in 2023 [29] and for the Australian carbon tax, which was repealed two years after its implementation. For some, more recently implemented, policies this may be explained by the time needed for sufficient data to become available, be assessed, and the results published. But even of the 38 carbon pricing schemes already implemented by 2015, for 18 of these we could not find a single study on effectiveness, despite the broad set of bibliographic databases searched. This confirms the evidence gaps identified in Chapter 2 for many of the implemented schemes. There is also little evidence on the effectiveness of carbon pricing relative to the level of the carbon price (carbon price elasticity). We identify only nine price elasticity studies, providing too few effect sizes for meta-analysing these separately.

Policy	**Jurisdiction**	**Introduction**	**Sector coverage**	**Emission coverage**	**Mean price**	**Studies**	**Effect sizes**
Chinese pilot ETS						46	179
o/w Hubei pilot ETS	Hubei, China	2014	industry	27%	$3	4	13
o/w Beijing pilot ETS	Beijing, China	2013	industry, power, transport and buildings	24%	$8	2	3
o/w Shanghai pilot ETS	Shanghai, China	2013	industry, buildings, transport	36%	$4	2	3
o/w Guangdong pilot ETS	Guangdong, China	2013	industry, aviation	40%	$5	2	2
o/w Shenzhen pilot ETS	Shenzhen, China	2013	industry, power, buildings, transport	30%	$7	1	2
o/w Tianjin pilot ETS	Tianjin, China	2013	industry, buildings	35%	$4	2	2
EU ETS	30 European countries	2005	power, manufacturing industry, aviation	38%	$20	13	77
Swedish carbon tax	Sweden	1991	transport, buildings	40%	$103	2	77
BC carbon tax	British Columbia, Canada	2008	industry, power, transport and buildings	70%	$18	7	39
Saitama ETS	Saitama, Japan	2011	industry, power, buildings	17%	$108	3	20
Tokyo ETS	Tokyo, Japan	2010	industry, power, buildings	20%	$106	4	14
Quebec ETS	Quebec, Canada	2013	industry, power, transport and buildings	77%	$9	2	10
RGGI	11 northeastern US states	2009	power	14%	$3	8	10
UK carbon price support	United Kingdom	2013	power	24%	$22	4	10
Finnish carbon tax	Finland	1990	industry, transport, buildings	36%	$6	2	8
Swiss ETS	Switzerland	2008	industry, power	11%	$18	1	5
Australian carbon tax	Australia	2012*	industry, power	60%	$24	1	2
California CaT	California, USA	2012	industry, power, transport, buildings	74%	$12	2	2
Korea ETS	Korea	2015	industry, power, buildings, domestic aviation, public sector, waste sector	74%	$15	2	2
Cross-country						4	18
Total						**101**	**483**

Table 3.1: **Carbon pricing schemes covered in the review**: All information on the carbon pricing schemes was retrieved from the World Bank [29], except for the price data for the EU ETS, which is retrieved from ICAP [201]. The information for the sector coverage was simplified. For more detailed information on the coverage, including covered or exempted subsectors, the reader is referred to the World Bank data. *Cross-country* studies analyse countries with and without carbon pricing, not focusing on a specific carbon pricing scheme. The effects of the eight Chinese pilot ETS schemes are often analysed collectively in a single study, while some studies focus on individual schemes. We only list pilots that have been studied individually. The Australian carbon tax was revoked in 2014. Mean prices are unweighted average prices in constant 2010 US$ during the period analysed by the studies in our sample. "Emission coverage" is the share of a jurisdictions emissions covered by the carbon price in 2022. The number of studies exceeds the number of reviewed articles, as some articles include more than one relevant study using disparate datasets.

3.3.2 Average emissions reductions across carbon pricing schemes

In order to provide a meaningful and transparent synthesis of the available quantitative evidence, we harmonise the effect sizes extracted from the individual studies to a common treatment effect, expressed as the percentage difference between the counterfactual emissions without carbon pricing and observed emissions after the introduction of a carbon price. It assumes emissions reductions to take place at the time of the introduction of the policy and to persist throughout the observation period as a constant difference to counterfactual emissions. Most of the reviewed studies report the emissions reduction effect as a treatment effect. Effect sizes provided as price elasticity are transformed into the same metric. Overall, we harmonise 483 effect sizes from 80 reviewed articles, covering 21 carbon pricing schemes that provide the starting point for our quantitative synthesis.

Our results show that carbon pricing effectively reduces greenhouse gas emissions. We use multilevel random and mixed effects models to account for dependencies among effect sizes in our sample and estimate the average treatment effects. The mixed effects model includes dummy variables for each of the included carbon pricing schemes to estimate the effectiveness for each of the schemes. As depicted in Panel a of Figure 3.2, emissions reduction effects are observed consistently across schemes with considerable variation in magnitude. For 17 of the carbon pricing schemes we find statistically significant average reduction effects from the introduction of a carbon price. The estimated reduction effects range from about -21% to about -5%. Across carbon pricing schemes, we find that on average the policy has reduced emissions by -10.4% [95% CI = (-11.9%, -8.9%)]. This effect is both substantial and highly statistically significant.

The reviewed literature provides large differences in the amount and quality of evidence for individual schemes. Focusing on those with the largest evidence base, we find an average treatment effect for the eight Chinese ETS pilots of -13.1% [95% CI = (-15.2%, -11.1%)], which is higher than the -10.4% average treatment effect across the schemes. The EU ETS and the British Columbia carbon tax both have estimated emission reduction effects below the overall average treatment effect. These are estimated at -7.3% [95% CI = (-10.5%, -4.0%)] and -5.4% [95% CI = (-9.6%,

-1.2%)]. Reduction effects smaller than -5% are only reported in three instances with severe problems in study design exposing estimates to a high risk of bias (Korean ETS, Australian carbon tax, Swiss ETS).

3.3.3 Critical appraisal and publication bias

The average treatment effects presented in the previous section were based on all reviewed studies. However, the quality of the primary studies is not uniform and some are subject to biases in the study design. Additionally, the average treatment effect might be subject to publication bias. Therefore we re-estimate the treatment effects by adjusting for potential quality issues and publication bias, adopting transparent and reproducible criteria.

We critically appraise each primary study, to identify potential biases in the study design. These biases often arise from the unreasonable selection of a control group used in a quasi-experimental design; from inadequately controlling for confounding factors like the introduction of other relevant policies; or from statistical specifications that do not allow to single out the policy effect. 46% of the reviewed studies are assessed to have a medium or high risk of bias. When we remove studies with medium or high risk of bias from the sample, the average treatment effects for some of the schemes are adjusted by up to 5 percentage points, while the estimation uncertainty increases due to the reduction of considered primary estimates (see Figure 3.2, Panel d). The identified biases, however, do not systematically impact the estimated treatment effects in either direction. The average treatment effect across policies is practically unchanged when removing studies with medium or high risk of bias.

Secondly, we adjust the average treatment effect for the influence of publication bias. Publication bias could arise from a tendency in the literature towards only publishing statistically significant effects [198, 209–211]. A precision effect test [195, 196] confirms the presence of publication bias in the set of studies reviewed here (see Appendix Table B.2). As suggested in the literature, we correct for publication bias by estimating average effects for a subsample of effect sizes with adequate statistical power [198], which applies to about 30% of the reviewed effect sizes. This subsample

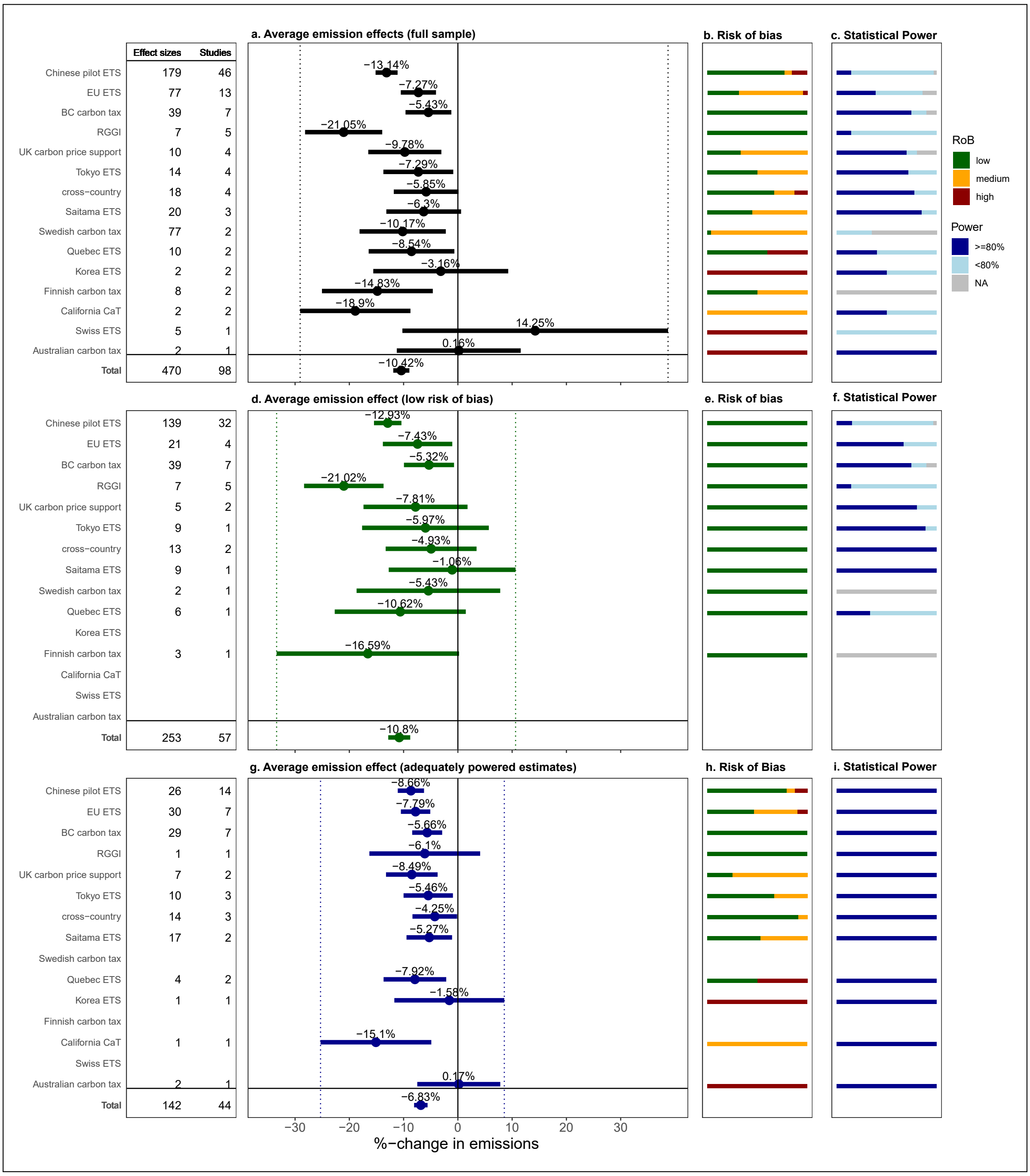

Figure 3.2: **Average emission changes by scheme**: The average effect sizes in panel a, d, and g are depicted together with their 95% confidence intervals based on multilevel random and mixed effects models and represent the effect of the policy observed in each period after its introduction in comparison to the counterfactual emissions without the policy. The estimates are ordered according to the number of studies they comprise (depicted on the left). The average treatment effect for the Chinese ETS pilots comprises the effects of all eight regional pilot schemes. *Cross-country* collects the evidence from studies assessing countries with and without carbon pricing, not focusing on a specific carbon pricing scheme. Panels b, e, and h show the distribution of assigned risks of bias. Panel c, f and i show the distribution of statistical power. Power above 80% is considered adequate. For synthetic control designs no statistical power was derived, thus presented as "NA". All estimates are also provided in the Appendix Tables B.2 and B.3.

analysis adjusts most of the scheme-wise average treatment effects towards lower estimated emissions reductions (see Figure 3.2, Panel g), ranging from -15% to -4%.

Across the schemes, the average treatment effect is reduced to -6.8% [95% CI = (-8.1%, -5.6%)]. Despite these adjustments, the publication bias corrected estimates support the overall finding that carbon pricing policies cause significant reductions in in greenhouse gas emissions.

Studies with a high risk of bias and low power are not uniformly distributed across schemes. Some schemes are evaluated only by a few biased studies, resulting in very high or low average treatment effects. For example, when considering all available evidence, the carbon pricing schemes in South Korea, Switzerland, and Australia are estimated to have the lowest negative or even positive average treatment effects. These estimates are based entirely on studies with a high risk of bias and are no longer considered when re-estimating the treatment effects based on low risk of bias studies (see Figure 3.2, Panel d). The two carbon pricing policies from the United States (California CaT, RGGI), which show the largest negative average treatment effect when considering all available studies, show lower average treatment effects after the adjustment for publication bias (see Figure 3.2, Panel g). For other schemes, like the EU ETS and British Columbia's carbon tax, there is no substantial change in the average treatment effect when studies with high risk of bias are excluded.

3.3.4 Explaining heterogeneity in effect sizes

There is considerable variation in the effect sizes reported by primary studies included in this review. This could arise from heterogeneity in the design of the carbon pricing policies or from heterogeneity in the design of the primary studies. The carbon pricing literature mainly discusses three policy design factors that could potentially explain differences in the effectiveness of the policy. First, there are debates whether carbon prices are better applied as carbon taxes or as emission trading schemes [165, 212–216]. Secondly, it is argued that the policy causes different reduction rates in different sectors [217–219]. And thirdly, the level of the carbon price can be expected to play a decisive role for the magnitude of the emission reductions [165, 220, 221]. We assess whether, and to what extent, such factors are able to explain differences in the treatment effects reported. We test which factors are most relevant to explain the reported emissions reductions by using scheme and study characteristics as explanatory variables in meta-regressions.

As we are confronted with a large number of potentially relevant explanatory variables, we use Bayesian model averaging (BMA) to assess the heterogeneity in the estimated effect sizes reported by the different studies. BMA is particularly suitable for meta-analysis as it allows for running a large number of meta-regressions with different possible combinations of explanatory variables and does not require selecting one individual specification. We include explanatory variables for the three policy design factors provided above: price level, sector coverage, and a variable differentiating between carbon taxes and cap-and-trade schemes. In addition we add dummy variables for each of the carbon pricing schemes, capturing the remaining policy design and contextual factors of each policy scheme. Additionally, we test whether studies assessing longer periods after the policy implementation find higher or lower treatment effects. To assess the impact of methodological choices made in the studies, we study a set of variables including the type of study design, estimation method, and data used in the primary studies.

The results from the BMA are provided in Table 3.2 and Figure 3.3. The posterior inclusion probability (PIP) indicates the relevance of each variable. Commonly, variables with a PIP above 0.5 are interpreted to be relevant explanatory factors, while variables with lower PIPs are unable to capture the observed heterogeneity. The table furthermore provides the posterior mean and standard deviation of the estimated effect averaged across all meta-regressions that include the respective variable.

Variation in carbon prices, the sectoral coverage of schemes, and choice of carbon tax vs. cap-and-trade do not seem to be important variables in explaining the observed heterogeneity in emissions reductions (PIP$<$0.5). Instead the dummy variables for the place where the schemes are applied do a better job in explaining this heterogeneity than the variables that capture specific design characteristics. The variables for the RGGI and the Chinese ETS pilots have a larger reduction effect on emissions than the EU ETS, which is set as the reference category. The Swiss ETS is estimated to have less of a reduction effect compared to the benchmark. Alternative specifications of the BMA, provided in Appendix Table B.5, also estimate a larger reduction effect for the Swedish carbon tax compared to the benchmark. The directions of these coefficients are in line with the average treatment effects presented in Figure 3.2, for the respective geographies.

If we remove the dummy variables for the schemes, the size of the carbon price becomes an important variable in the BMA to explain the heterogeneity in emission reductions with a PIP close to 1 (see Appendix Table B.6). However, in the absence of the scheme dummies the effect of the price variable is likely to be confounded as the scheme dummies account for any omitted context variable that does not vary within a scheme. The high correlation of 0.96 between the scheme dummies and the price variable indicates that the price variable captures the heterogeneity between schemes. In fact, the price coefficient is estimated with a positive sign in the BMA specification without the scheme dummies, implying that lower emission reductions are achieved with higher carbon prices. The counterintuitive direction of the price effect indicates a misspecification of the model when the scheme dummies are excluded. Below we discuss possible causes for this inverse relationship between the price and the reduction effect in our data. The effect of carbon prices on emissions reductions is better identified by adding scheme dummies to focus on the variation of prices within each scheme. However, the largest share of the variation in our carbon price variable comes from variation between the schemes (91%) and only 9% from within scheme variation. This is not a limitation of our dataset. Indeed, carbon prices tend to vary strongly across countries based on the design and coverage of scheme. But for individual schemes prices have historically been stagnant (EU ETS till recently, RGGI, Chinese ETS pilots) or increases relatively modest (British Columbia carbon tax) [201] and the effect size estimates evaluated here provide limited time frequency. We suspect that due to this low variation, our sample has insufficient power to identify carbon prices as a relevant factor in explaining emission reductions.

Studies assessing the effectiveness of carbon pricing over longer time periods find larger emission reductions. The coefficient for the variable *duration* has a PIP of 0.76 and is estimated with a negative sign for all regression specifications it is included in. Testing for the spatial and temporal granularity of the data suggests that only the use of city level data compared to the country level explains some of the heterogeneity in reported effect sizes. Methodological differences in the reviewed studies only have a minor influence on effect sizes.

In line with the previous section, we also include the risk of bias variable and the standard error, capturing the publication bias. They are both not detected to be most relevant to explain the heterogeneity.

3.4 Discussion

In this first quantitative meta-analysis of carbon pricing evaluations, we find robust evidence that existing carbon pricing schemes that have been evaluated to date are effective in reducing greenhouse gas emissions. Our machine-learning enhanced approach to study identification finds more than twice as many ex-post evaluations than existing reviews [18, 137–139], studying the effectiveness of 21 carbon pricing policies. Our meta-analysis finds that at least 17 of these policies have caused significant emissions reductions ranging from -5% to -21%. These are substantially larger than the 0% to -2% suggested in the recent and widely cited review by Green [137], which lacks a clear and transparent methodology to synthesise the literature [19], not allowing us to formally compare our results. Our finding is robust to biases from poor study designs as well as publication bias. Correcting for the latter adjusts the range of observed emissions reductions to -4% to -15% across carbon pricing schemes.

The synthesis of research findings across carbon pricing schemes provides comprehensive and consistent evidence of its effectiveness, despite the heterogeneity of policy designs and regional contexts. Compared to the recent assessment report by the IPCC, which provides a quantification of achieved reductions only for the EU ETS [9], our systematic review adds synthesised emission reduction estimates for more than a dozen carbon pricing schemes. We provide these estimates together with uncertainty ranges and a transparent assessment of study quality and highlight the presence of substantial variation in emission reductions achieved across the schemes in our sample, ranging from -5% for the carbon tax in British Columbia to -21% for the RGGI. We conduct an early application of Bayesian model averaging for meta-regressions on our dataset of 483 effect sizes to disentangle which factors explain

	PIP	Post Mean	Post SD
RGGI	**1.00**	**-28.45**	**5.09**
Chinese_pilot_ETS	**0.99**	**-9.76**	**2.23**
Swiss_ETS	**0.80**	**14.35**	**8.93**
Data_City	**0.78**	**11.39**	**7.63**
duration	**0.76**	**-0.64**	**0.46**
synthetic_control	0.42	2.87	3.87
tax	0.41	-3.11	4.24
BC_carbon_tax	0.38	3.90	5.65
Swedish_carbon_tax	0.36	-3.05	4.65
Coal	0.32	-2.58	4.26
Less_Bias	0.30	1.16	2.00
Finnish_carbon_tax	0.25	-2.89	5.70
TransLevelLevel	0.19	-0.70	1.67
Data_Region	0.12	-0.40	1.30
log_carbon_price	0.09	0.15	0.62
Data_Sector	0.09	0.25	1.02
Gas	0.08	-0.44	1.88
other_schemes	0.05	-0.31	2.00
Tokyo_ETS	0.04	0.15	1.03
industrial_sectors	0.04	-0.04	0.77
Data_Firm	0.04	0.07	0.54
Data_Plant	0.03	0.04	0.47
DVTotal	0.03	0.03	0.51
Saitama_ETS	0.03	0.05	0.65
SE_percent	0.03	-0.00	0.00
Gasoline	0.03	-0.02	0.62
Quebec_ETS	0.03	-0.06	0.82
Data_Month	0.03	-0.02	0.52
Data_Year	0.03	0.01	0.41
Data_Airline	0.03	-0.00	0.75
(Intercept)	1.00	-5.99	

Table 3.2: **Heterogeneity assessment using Bayesian model averaging:** The table provides the results of meta-regressions using Bayesian model averaging. The dependent variable for each of the meta-regression models is the percentage change in emissions. The posterior inclusion probability (PIP) indicates the relevance of each variable. Variables with PIP≥0.5 are considered relevant for explaining the heterogeneity in carbon emissions reductions reported across primary studies. Post Mean and Post SD represent the mean and standard deviation of the posterior distribution for a respective explanatory variable. Five variables have PIP≥0.5 and are considered relevant: the dummy variables for *RGGI*, *Chinese pilot ETS*, *Swiss ETS*, *Data_City*, and *duration*. The dummy variables represent the geographic location in which the policy was implemented, with the reference location being EU ETS. *Data_City* captures whether primary studies used city level data versus country level data. The variable *duration* captures the number of years for which data on the scheme was collected after the policy was implemented. Definitions of the other explanatory variables are provided in Appendix Table B.4.

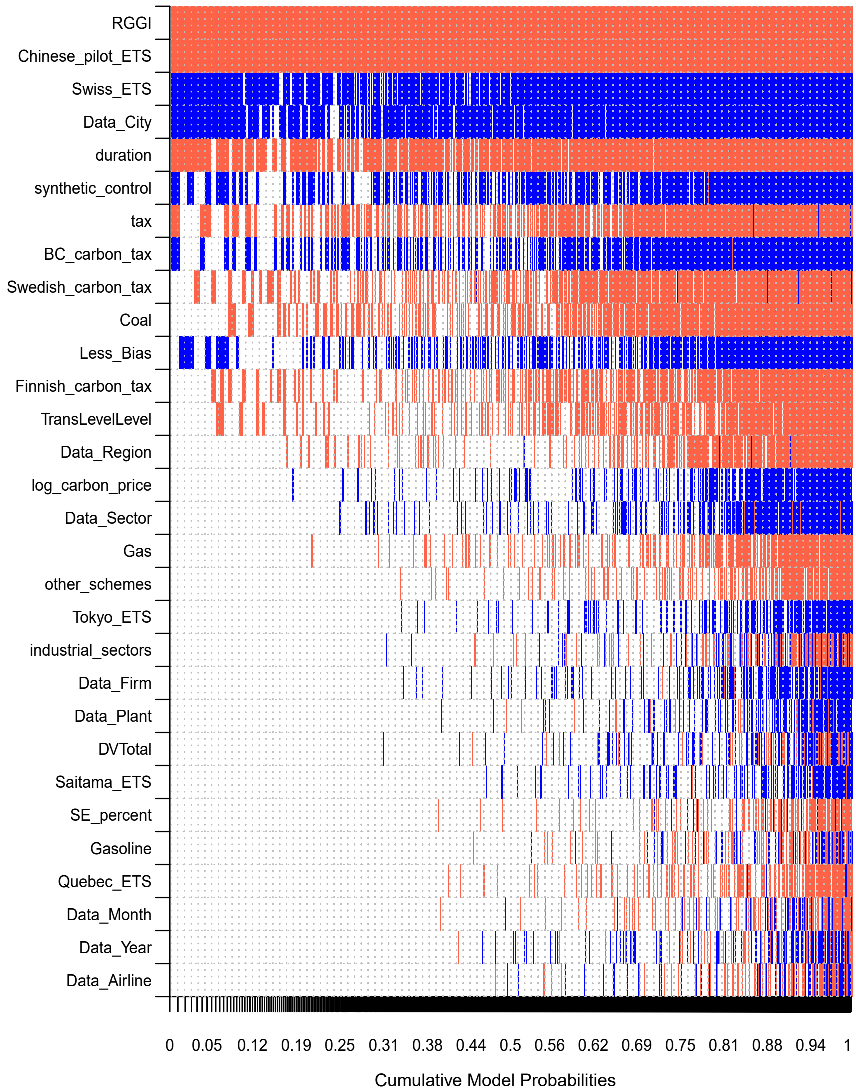

Figure 3.3: **Heterogeneity assessment using Bayesian model averaging**: The columns in the figure depict the best 26,435 estimated meta-regressions, with each column showing the outcome of one estimated meta-regression model. The dependent variable for each of the meta-regression models is the percentage change in emissions. The possible explanatory variables are depicted in the rows (ordered by their PIP in descending order) and the explanatory variables included in a respective meta-regression model of the column is indicated by the colours. Red colour indicates the variable was included with a negative sign (larger emission reductions). Blue colour indicates a positive sign (smaller emission reductions). No colour indicates that the variable was not included in the meta-regression model represented by that column. The horizontal axis indicates the cumulative posterior model probabilities across all models. The models are ranked by their posterior model probability with the model on the left accounting for the largest posterior model probability. The definitions of the explanatory variables are provided in Appendix Table B.4.

these differences. The findings suggest that the individual context and policy design of the schemes best explain the heterogeneity in achieved emissions reductions. These are the most relevant explanatory factors despite controlling for broader policy design features like the sectoral coverage or the design as carbon tax or carbon trading scheme as well as for study design features of the primary studies.

Our heterogeneity analysis does not identify a relationship between the price level and the achieved emissions reductions, i.e. the size of the emissions reductions observed across schemes from the introduction of a carbon price cannot be explained well by the carbon price level. This is not surprising as marginal abatement costs may differ widely as, for example, prominently acknowledged in the literature on linking carbon pricing schemes [222, 223]. It is further different from the expectation that higher carbon prices lead to larger emissions reductions within a carbon pricing scheme as commonly found in available assessments of fuel price elasticities [183, 224, 225]. In line with this argument, we find that the relationship between carbon price levels and emissions reductions in our meta-analytic framework is dominated by the across-scheme variation in prices, which accounts for 91% of the variation in our dataset while the variation within schemes only accounts for 9%. The interpretation for not finding a clear relationship should thus rather be that when implementing a carbon price in two countries with different country contexts, the country with the higher carbon price would not necessarily experience the higher emissions reductions.

This can be observed, for instance, when looking at the cases of China, the EU, and British Columbia. The reviewed literature finds larger emissions reduction effects for the pilot emission trading schemes in China (-13.1%) than for the EU ETS (-7.3%) and the carbon tax in British Columbia (-5.4%), despite the very low carbon prices of the Chinese schemes. The average prices of the eight Chinese pilot schemes are all below US$ 8 during the study period, while the average prices for the EU and British Columbia are at US$ 20 and US$ 18, respectively. This is likely a result of lower abatement costs in China [226] together with differences in the policy contexts of the countries. The effectiveness is certainly influenced by other policies in place. In China indirect carbon prices are lower than in the EU countries and Canada [227], allowing for a higher marginal effect of the implementation of the ETS pilots in China. Non-pricing instruments also diverge across countries. In addition, the implementation of a carbon price (even with a low price) can have a signalling effect towards the emitters, underlining the commitment of the government towards climate mitigation. Evidence for the Guangdong province suggests that signalling has significantly contributed to the achieved emissions reductions in the context of the introduction of the ETS pilots in China [228]. Another example highlighting the relevance of the context of the policy implementation is the case of the RGGI. The policy implementation coincides with the shale gas boom, which drastically reduced the prices of natural gas in the USA and started around the same time as the RGGI was implemented. In face of these general price dynamics in the US energy sector, RGGI participating states reduced their emissions considerably stronger compared to non-regulated states [229, 230], while the carbon price was only US$ 3 on average.

Even if across schemes the price level of the carbon price, is not found to be the relevant driver of the emissions reductions achieved with the introduction of the policy, within a scheme the effectiveness is expected to increase with increasing prices. This is well studied for other changes in fuel prices, which are found to substantially reduce its consumption [224, 225]. That literature studies all possible price changes on a single fuel, while the here assessed literature on carbon prices studies the effect of a single policy instrument across all fuels. It is thus a complementary but distinct body of evidence. Meta-analyses estimate a reduction of fuel consumption between 0.31% and 0.85% in the long run for a 1% increase in the fuel price [179–183].

Within the literature evaluating the policy effectiveness we identified only nine primary studies estimating semi-elasticities of carbon prices. Four are using the step-wise introduction of the carbon tax in British Columbia to estimate elasticities for the transport and buildings sectors [176, 177, 231, 232], while one is conducted respectively for RGGI [230] and EU ETS [178]. In addition, some studies estimate elasticities across countries and carbon pricing schemes [18, 233, 234]. These studies support what was already known from studies on the price elasticity of fuel consumption [179–183, 224, 225]: increasing prices reduce fuel use and emissions. Hence, as carbon prices further rise after the introduction additional emissions reductions are achieved. Interestingly, some studies suggest that an increase in the carbon tax leads to larger emissions reductions than an increase of the same size in the market price of the fuel [176, 177, 231, 232]. It will thus be a relevant avenue for future research to understand whether it is a generalisable finding that price elasticities are higher for policy induced price changes compared to market price changes of fossil fuels. Such research could draw on the comprehensive evidence from the fuel price literature.

Our meta-regression results suggest that the policy effectiveness of carbon pricing policies increases with time. Studies covering longer time periods after the introduction of the carbon price report larger emissions reduction effects compared to assessments for shorter time periods. While this finding should be treated with caution, as most of the primary studies assume constant treatment effects for their estimations, it hints towards increasing emissions reductions in the years following the policy introduction. The assumption of constant treatment effects reflects not only methodological considerations of the primary studies, but is also based on the expectation that as long as the carbon price of the implemented policy is unchanged, the emission reduction effect should not intensify. The finding of our meta-regression to some extend counters that assumption. An increasing policy effectiveness could be a result of steady adjustment processes, enforced by innovation and investments into cleaner production and infrastructure. Additionally, the literature reviewed here provides some evidence that an increasing policy stringency has also played its role in strengthening the effectiveness of the policy. Increases in the carbon prices led to additional emission reductions in Sweden [235] and the United Kingdom [236]. Similar effects are found for the EU ETS, where the effectiveness increases with the increasing stringency from phases I, II, and III [237–240].

While the harmonisation and synthesis of the emissions reduction effects provides an overview of the policy effectiveness across a large number of policy schemes, it raises a number of policy relevant research questions, which cannot be answered with a purely quantitative, meta-econometric approach – which is inherently dependent on the available evidence base. These limitations could be addressed using more configurative review designs such as realist synthesis which systematically combine complementary lines of evidence to better understand why particular policy designs work, under what conditions, and why [6, 28]. This is done in Chapter 4. Some research gaps, however, need to be filled by further primary research. First, there are more than 50 carbon pricing schemes that have not yet been evaluated for their emissions reduction effect, despite some of them being enacted for more then ten years. Others have still been studied insufficiently or only poorly. Second, we lack ex-post evidence of higher carbon prices. There are currently less than ten studies assessing emissions reductions in schemes with mean carbon prices higher than US$ 30 across the observation period. As policy ambitions are raised over time, there is an opportunity to strengthen that evidence base. Thirdly, this systematic review highlights substantial challenges with potential biases in the available primary evidence. Only about half of the studies assessed here follow rigorous study designs with a low risk of bias and only 30% of the studies are adequately powered. While some of this might be related to a lack of access to adequate data for the most rigorous research designs, high quality primary research is essential to understand the effectiveness of climate policies [241]. The multitude of supplementary or conflicting policies as well as other confounding factors pose a challenge to the clear identification of the causal effects of a specific policy [242]. Novel methods of reverse causality are a promising avenue to address this challenge [243].

The effectiveness is just one dimension of policy outcome relevant to the selection of the best policy measures. Systematic assessments of the ex-post climate policy literature on a multitude of policy outcomes and different climate policy options could be the basis for accelerated learning on climate policies and considerably improve upcoming IPCC assessments. Unless we raise our standards and do this work, policymakers and society will remain in the dark as to the most promising pathways towards addressing the climate crisis.

Chapter 4

Realist synthesis: Under what conditions and why is carbon pricing effective?

4.1 Introduction

Thousands of implemented climate policies provide an excellent environment to learn from their success and failure [12, 13], which can inform policymakers to get on track towards net-zero emissions. Previous studies, including the meta-analysis in Chapter 3, have shown that the effectiveness of climate policies varies considerably across instruments and contexts [16, 18]. To benefit from the rich amount of policy experiences, gathered around the world, we need to move beyond the question what has been effective or not, and study in more detail under what conditions and why a policy is or is not effective. A nuanced understanding of the context factors that make a policy more or less successful are crucial to form expectations to what extent policy experiences can be transferred from one country to another.

Evidence synthesis can play a key role in fostering cross-country policy learning. There are large amounts of country-specific policy evaluations already available [15] (see also Chapter 2). These contain a treasure of detailed evidence, what it is about a policy that makes it more or less effective in a given context. The array of evidence synthesis methods, listed in Table 1.1, provides a glimpse of the variety of synthesis methods, capable of capturing and synthesising different nuances from heterogenous types of evidence, required for a context-sensitive assessment of the policy experiences [27, 50]. We need to make best use of the complementary synthesis methods to exploit the full potential of the rich evidence, provided by thousands of policy evaluations.

The meta-analysis, conducted in Chapter 3, provides robust evidence that implemented carbon pricing policies have achieved significant emissions reductions, but provides limited insights to explain the heterogeneity in reduction rates observed across contexts. It finds that the design of the policy as a carbon tax or cap-and-trade scheme, the level of the carbon price, and the sector the policy is applied to can all not explain the cross-country variation. Instead, the meta-regression results suggest that the explanations need to be found in the specific country-contexts of the policies. This needs to be studied, drawing on further lines of evidence.

In this chapter, we use a realist synthesis design, to study what it is about carbon pricing policies, that makes it more or less effective across contexts. Realist synthesis systematically assesses the causal mechanisms triggered by an intervention and evaluates what context factors influence the mechanism to unfold. It iteratively collects hypotheses on combinations of mechanisms and contexts and triangulates heterogenous lines of evidence to test the validity of the hypothesis. In this way, the synthesis generates an evidence-based theory, under what conditions and why the policy is more or less effective [6, 28].

To our knowledge realist synthesis has not yet been applied to the field of climate policy evaluation. Berrang-Ford *et al.* [244] explore the applicability of the method to climate change adaptation practices and Sawatzky *et al.* [245] and Muzorewa and Chitakira [246] provide the first applications in the field of climate change adaptation. Sarmiento Barletti *et al.* [247] use realist synthesis to study community engagement practices in the land-use context. In addition, a number of studies have applied realist synthesis to study health outcomes of energy efficiency interventions in household heating [127–129]. We here adopt and refine the method to synthesise the evidence from climate policy evaluations and evaluate the applicability of the method for climate policy research.

In this chapter, we aim to compile evidence based theories on the working mechanisms of carbon pricing across schemes and sectors. We use a realist synthesis methodology to answer three interrelated research questions: (1) What are the hypothesised mechanisms by which carbon pricing schemes are effective? (2) What is the role of different contexts in moderating these mechanisms? (3) What is the underlying evidence supporting these hypotheses? Our dataset is a compilation of 80 carbon pricing evaluations, obtained through a systematic literature search and screening process. From this we extract 177 hypotheses on how and under what conditions the policy works and synthesise these into nine main hypotheses. Based on 293 pieces of evidence on 21 carbon pricing schemes, we test and refine the theories as a foundation for a more faceted theory of how carbon pricing leads to emissions reductions.

4.2 Realist synthesis

Realist synthesis aims to derive a nuanced understanding on which policy is suitable in which context. It studies who is supposed to react to the policy in which way and what are the preconditions to achieve these reactions, building on a variety of quantitive and qualitative lines of evidence [6, 28]. While effectiveness studies, like the meta-analysis conducted in Chapter 3, put their emphasis on the ultimate outcome of the policy intervention, assuming a causal link between policy and outcome, realist approaches aim to unpack this causal link [28]. What are the mechanisms triggered by the policy leading to the outcome? Where are potential blockages hindering the mechanism to unfold? And how do different contextual settings alter the causal linkages? Realist synthesis aims to answer these questions by searching for and evaluating available evidence.

The synthesis is guided by the development and testing of hypotheses and theories. In the realist framework these are expressed as context-mechanism-outcome (CMO) configurations, as depicted in Figure 4.1. In a given context C, the policy triggers a mechanism M, which leads to an outcome O [28, 248]. As an example, a feed-

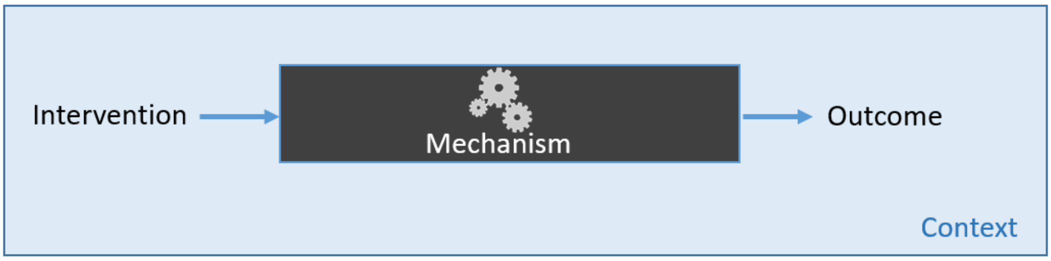

Figure 4.1: **Context-mechanism-outcome framework**: Realist synthesis aims to unpack the black box, between the implementation of an intervention and its outcomes. This is conceptualised in a framework, considering contexts and mechanisms relevant to cause a given outcome [6].

in tariff for photovoltaics could trigger homeowners with a south-facing roof (C) to seek consultation on the suitability and profitability of their location (M), leading to the installation of a solar panel (O). Realist synthesis aims to collect and synthesise theories on the mechanisms and context factors leading from a policy intervention to its outcome(s). These should be specified as testable hypotheses for the further investigation [249].

The collected hypotheses are tested against the best available evidence, identified through systematic and transparent searches of the literature. Unlike other synthesis methods, such as for instance meta-analysis, realist synthesis is not restricted to evidence in pre-defined formats [6]. Instead it triangulates heterogenous evidence to test the validity of a theory. The reviewed evidence may also suggest amendments to the theories, for example, where conflicting findings point towards relevant, previously unconsidered, context factors that may explain why a mechanism was triggered in some but not all reviewed cases. In this way, realist synthesis aims to refine our understanding of the policy rather than quantifying or ranking the relevance of different mechanisms [6, 250].

Realist synthesis follows an iterative approach of searching for hypotheses and testing these against the available evidence [6]. The broad questions asked by realist syntheses open the floor for a wide range of potential theories, combined with the appreciation of various forms of evidence, this could lead to infinite iterations of refining and testing hypotheses. Realist synthesis will, therefore, not be able to review relevant evidence exhaustively [28]. While this distinguishes this approach from more aggregative reviews, studying a narrow research question with the aim to capture all evidence on that question [50], it does not relieve the reviewer of a transparent and rigorous assessment of the primary evidence [6, 250].

4.3 Methods

Given the observed differences in the effectiveness of carbon pricing schemes across the globe, identified in Chapter 3, we use realist synthesis to develop theories of mechanisms how carbon pricing causes emissions reductions in various contexts. Carbon pricing is commonly advocated for its simplicity. It charges a uniform price for the emission of each ton of carbon dioxide or its equivalents – ideally across sectors. Its core assumption is that policy makers do not need to worry about the specific mechanisms where and how emissions can best be reduced, but that all emitters are incentivised to reduce emissions to avoid the cost [165, 251]. We aim to refine this theory by studying the mechanisms where and how emissions are reduced, to gain a better understanding what it is about the policy that makes it more or less effective in different contexts.

4.3.1 Literature search

We synthesise the evidence from a set of available ex-post policy evaluations studying the emissions reduction effect of carbon pricing policies. We choose to review ex-post policy evaluations for three reasons. First, all reviewed studies assess a carbon pricing intervention and an emissions outcome. Secondly, the use of causal inference methods in the primary studies allows for the identification of causal mechanisms. And thirdly, studies applying causal inference methods on non-experimental data commonly engage intensively with the data to ensure their finding is not an artefact of any confounding factor, providing valuable information on mechanisms and contexts in the form of auxiliary assessments or the collection of supplementary data.

We use the same set of studies collected for Chapter 3. The systematic search and selection of studies is depicted in Figure 4.2. A broad set of search terms was used on four bibliographic databases (Web of Science, Scopus, JSTOR, RePEc) and Google Scholar, identifying more than 16,000 potentially relevant studies. After consecutively screening titles and abstracts followed by full texts, 80 relevant ex-post policy evaluations are identified, covering 21 carbon pricing schemes. A list of the included studies is provided in Appendix Table B.1.

4.3.2 Data extraction

In the data extraction stage, we inductively extract hypothesises on the working mechanisms of carbon pricing from the identified literature. The data extraction for the realist synthesis is conducted independently of the data that was extracted for the meta-analysis due to the different data requirements and methods of this study. We use the context-mechanism-outcome structure as a framework and code hypotheses and evidence on contexts and mechanisms from each study. We develop the codes used to classify a specific context or mechanism iteratively while coding. Hypotheses are often provided in the motivation of a study, in the methods section as research hypotheses, or in the discussion or conclusion section to provide possible explanations for unexpected research outcomes. Evidence is often provided in the results section. The coding therefore takes into account relevant information from all sections of each research paper. Each segment of the document, that provides a hypothesis or evidence on a context or a mechanism is extracted and labelled with all relevant codes. We assign labels from three categories: *context*, *mechanism*, and a distinction between *hypotheses* and *evidence*. If a new context or mechanism type is identified, a new label is created. The created label is used in the subsequent data extraction to classify text elements on the same topic. In that way, we created 28 context labels and 17 mechanism labels (see Appendix Table C.2). The context labels are amended by the 21 carbon pricing schemes in the sample, which are labelled on the study level.

Each hypothesis or evidence element is ideally assigned at least one context and one mechanism label, to cluster these in context-mechanism-outcome configurations. We assign the context and mechanism labels to the best of our knowledge from the information provided in the study. For some elements we have very specific context information available, allowing to assign a range of context labels, while for others the context cannot be specified further than the studied scheme. Similarly, some elements provide a detailed assessment of a specific mechanisms, while others are more

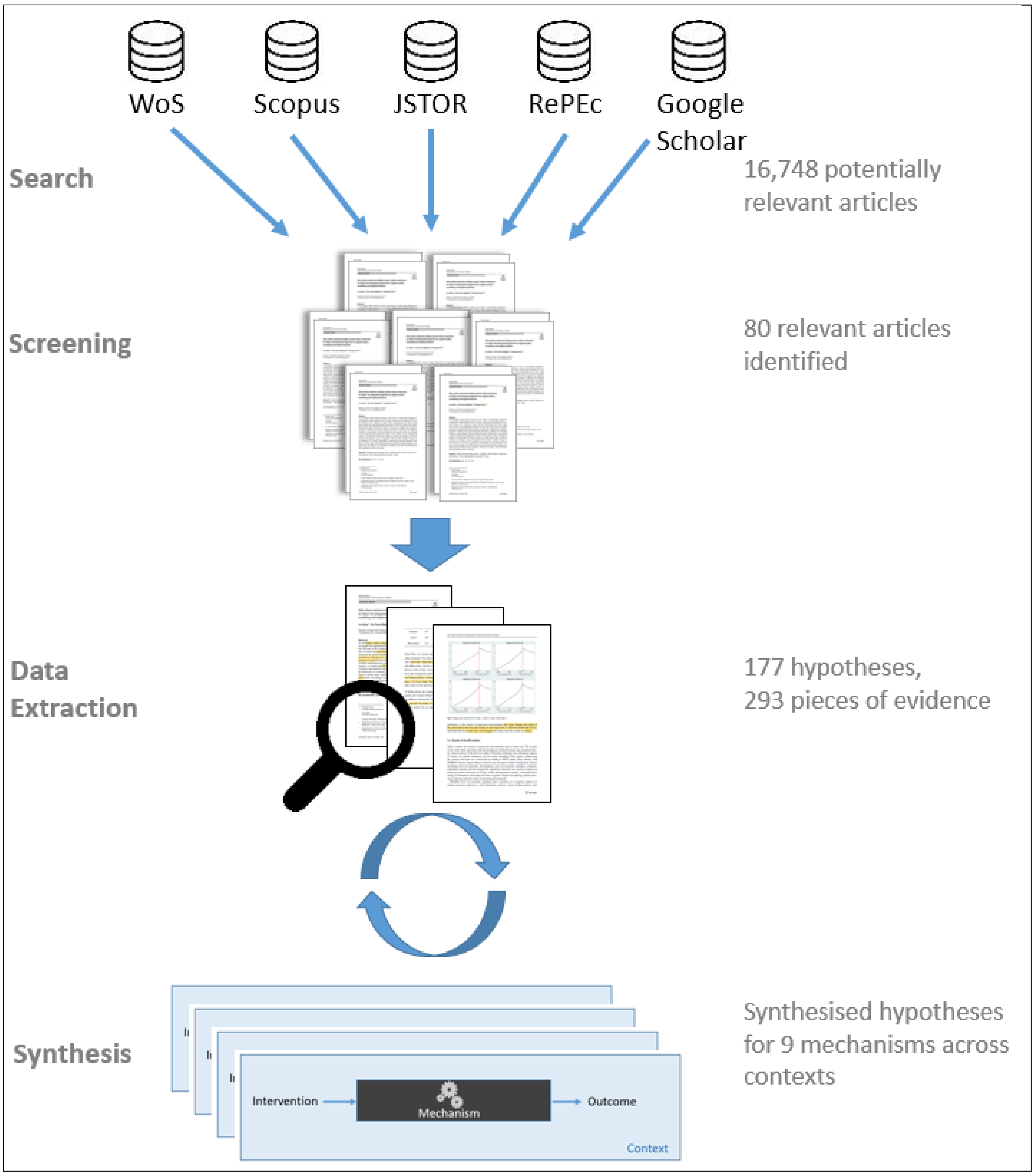

Figure 4.2: **Flow diagram of the systematic review process**: We searched for evidence in four bibliographic databases and a web based search engine. After screening we retained 80 relevant articles, from which we extract hypotheses and evidence on the mechanisms and contexts of carbon pricing. These are synthesised using the context-mechanism-outcome framework.

vague about the causal mechanism they assume. The context-mechanism-outcome information, therefore, varies in the level of detail assigned for each hypothesis and evidence element.

4.3.3 Synthesis

To synthesise the extracted data, we assess the labelled information – category by category – to identify common themes. We start by reviewing the identified hypotheses clustered by the assigned mechanism and context labels. We group the hypotheses extracted from the primary studies based on the mechanism they relate to. We also tested an alternative grouping based on the context categories, but find the mechanism structure to provide a more consistent framework for the synthesis.

For each group of hypotheses we formulate a main hypothesis describing the assumed mechanism. We then synthesise the hypotheses in each group to collect all assumptions how the mechanism is expected to unfold. In particular, we capture the context factors assumed to be relevant in explaining the occurrence and intensity of a given mechanism. This is facilitated by the context labels assigned to each hypothesis extracted from the primary studies. The level of detail provided for the context factors ranges from broad assumptions on sector differences to much more detailed assumptions what it is about a sector, firm, or household that influences whether a mechanism is triggered.

For each of the synthesised hypotheses we assess the availability of evidence to evaluate the hypothesis. We use the labels assigned during the data extraction to identify evidence relevant to each hypothesis.

The largest part of the extracted evidence is derived using statistical hypothesis testing. We evaluate such evidence in line with the reported significance test. Insignificant findings are interpreted as *no evidence*. For other types of evidence, i.e. evidence from descriptive assessments of data, we report how the evidence was derived and generally interpret it as suggestive evidence. While we do not conduct a comprehensive risk of bias assessment, we critically evaluate the reliability of the cited evidence and report caveats as required.

We report our findings in the results section below, structured by the mechanisms. For each mechanism, we first provide the synthesised hypothesis and highlight relevant details of expected variations based on the context. We continue by reporting on the evidence used to test the hypothesis, assessing to what degree the evidence supports the hypothesis. This is followed, for each mechanism, by a conclusion, which refines the hypothesis based on the assessed evidence.

4.4 Results

From 80 studies we extract 177 instances where hypotheses are mentioned, which we cluster by the nine main mechanisms (see Figure 4.2). We evaluate these hypotheses based on the evidence provided in the same 80 studies – namely a set of 293 statements. Most of the evidence results from quantitative quasi-experimental research methods, providing causal assessments of the hypotheses. It further includes evidence from surveys and descriptive assessments of data.

The nine mechanisms, derived from the synthesis of the hypotheses, comprise fuel switching, adoption of low carbon technologies, efficiency improvements, investment changes, research and development, reduced production and consumption, leakage, salience, and the anticipation of future carbon prices (see Figure 4.3)[1]. Hypotheses expressed in primary studies that could not be captured, as they could not be allocated to a specific mechanism, are provided in Appendix Table C.1.

The mechanism contexts expressed in each hypotheses 4.4.1-4.4.9 are mainly characterised by the policy scheme (including country of implementation and aspects of policy design) and the studied sector. The reviewed studies cover 21 carbon pricing schemes from various world regions, applied at the sub-national, national or regional level. In particular, we assess evidence from sub-national emission trading schemes (ETS) in eight Chinese provinces (commonly assessed together as a single policy), as well as in Tokyo, Saitama, Quebec, California, and a group of eleven US states (regional greenhouse gas initiative, RGGI), from a regional carbon tax

[1] Interestingly, a number of the mechanisms studied here are captured as policy outcomes in the evidence and gap map in Chapter 2. This indicates that some policy outcomes are not only relevant in their own right, but are, at the same time, relevant to explain the causal mechanism towards emissions reductions.

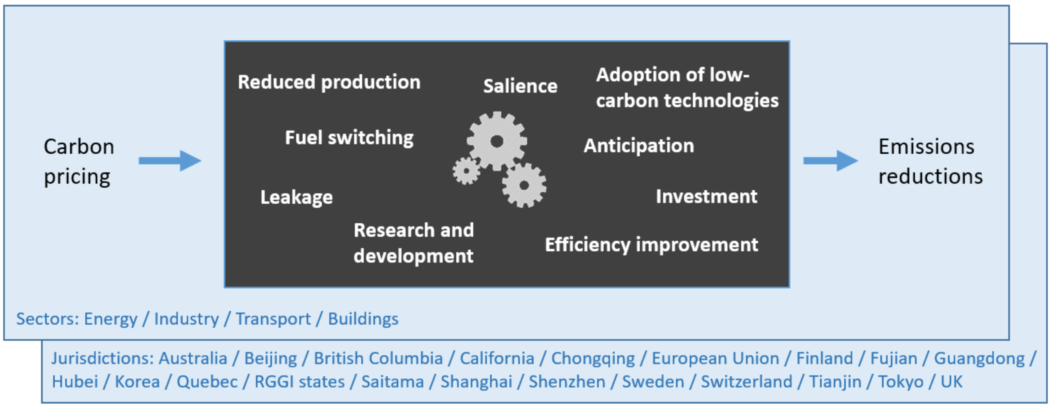

Figure 4.3: **Synthesis of hypotheses on nine mechanisms across contexts**: We group the identified hypotheses by the nine mechanisms by which carbon pricing is assumed to cause emissions reductions. The occurrence of these mechanisms depends on the context in the given sector and jurisdiction. The depiction of the context factors is simplified here and the details are provided in sections 4.4.1-4.4.9.

in British Columbia, national ETSs in Australia, Korea and Switzerland, national carbon taxes in Finland and Sweden, a regional ETS in the European Union (EU), and a national policy applied in the UK to increase the stringency of the EU ETS on national level (carbon price support). The sectoral context is synthesised using the IPCC emission sectors: energy, industry, buildings, and transport (with other emission sectors not being relevant). The reviewed policy evaluations study the policy for periods of one to 16 years of its implementation, with 80% of the studies assessing 7 years or less, thus focussing on short-run rather than long-term mechanisms. The mean price level across the studies is at about US$ 19 per ton of emitted carbon dioxide (equivalent).

An overview of the hypotheses and the associated evidence is provided in Table 4.1. The overview also points to gaps in the evidence, where a context-mechanism configuration could not be tested.

Table 4.1: **Mechanisms how carbon pricing is assumed to cause emissions reductions**: Summary of hypothesised mechanisms and reviewed evidence.

Mechanism	Explanation	Context	Evidence				
			Energy	Industry	Transport	Buildings	Across sectors
Fuel switching	Carbon pricing induces firms to substitute high emitting fuels for lower emitting fuels	Dependent on substitutable fuels and processes, and therefore may not apply to certain sectors such as aviation; In the short-term dependent on available capacities	Coal to gas transitions observed in the UK power sector [175, 178, 252, 253]; No absolute (potentially relative) fuel switching observed for EU ETS [178, 240, 254, 255] or RGGI [230, 256, 257]	Not observed for Tokyo or Saitama ETS [258]	Indications from Sweden and Finland for switching from gasoline to diesel cars [259, 260]	Indications from Sweden for switching from oil heating to wood, electricity, and district heating [235]	Relative rather than absolute fuel switching caused by Chinese ETS pilots [261–269]
Low-carbon technology adoption	Carbon pricing induces firms to substitute fossil fuels for renewable and low-carbon sources (e.g. wind, solar, hydro, nuclear-power, electricity)	Expected to be a more gradual and longer-term outcome, due to the lead-time for developing alternative energy sources	No uptake in renewable energies in Tokyo [258] and RGGI states [257]; UK carbon price increased hydro but not nuclear power [252]			Indications for increase in heat pump installations in Sweden [235]	

Table 4.1 (continued)

Mechanism	Explanation	Context	Evidence				
			Energy	Industry	Transport	Buildings	Across sectors
Efficiency improvements	Carbon pricing induces firms to adjust processes, to lower the emission intensity of their production	Costs for efficiency improvements differ by sector and firm; Adjustments are expected to require some time to be implemented	Increase in emission intensity observed in RGGI states [257]	EU ETS caused significant reductions in emission intensity in France [254, 255], Germany [240], and Lithuania [270], but not in Norway [271]			Chinese ETS pilots cause significant reductions in emission intensity [265, 267, 269, 272–274] (few studies find no such effect [275, 276]); emission intensity reductions vary by ETS pilot scheme [265]; both energy intensity [174, 272, 273] and emission intensity of energy reduced [269]; Emission intensity reductions observed in Quebec [277]
Investments	Carbon pricing incentivises firms to invest into new equipment to reduce emissions	Investment decisions are based on long-term expectations of carbon price levels and depend on financing conditions		EU ETS fosters investments of French manufacturing sector into reduced emission intensity [255]		Indication for increased investments into heat pumps fostered by Swedish carbon tax [235]	Chinese ETS pilots reverse effect of foreign direct investment from increasing to decreasing effect on emissions [278]

Table 4.1 (continued)

Mechanism	Explanation	Context	Evidence				
			Energy	Industry	Transport	Buildings	Across sectors
Research and development	Carbon pricing fosters research and development of new technologies to reduce emissions						Chinese ETS pilots increase expenditure for research and development [279] (another study does not find the same effect [261]); number of patents [263] and of "green" patents [268]; increased innovation is found to reduce emissions [263, 264]
Reduced production/consumption	Emissions can be avoided by reducing the production or consumption	Energy and industry sectors may be more affected, due to inelastic demand for heating and transport	Reduced energy generation caused by Chinese ETS pilots [174, 252, 262, 269, 273, 280] but not by Tokyo ETS [280]	No reductions of industrial output found for Chinese ETS pilots [281, 282] or EU ETS [240]	Reductions in short-haul flights (EU ETS) [283] and urban car use (British Columbia carbon tax) [231]		No evidence for reduced production caused by Chinese ETS pilots [265, 269, 281, 282, 284], British Columbia carbon tax [285], or European carbon taxes [286]

Table 4.1 (continued)

Mechanism	Explanation	Context	Evidence				
			Energy	Industry	Transport	Buildings	Across sectors
Leakage	Production and associated emissions are relocated to jurisdictions where they are not subject to the policy or reallocated to entities that are exempted from the policy	Dependent on spare production capacities outside the treated jurisdiction and, in the case of the energy sector, spare transmission capacities; Leakage is mainly assumed for energy and industry sectors (and to a limited extent in the transport sector)	RGGI and UK carbon price support cause increase in (lower emission) energy imports [230, 252, 256, 257]; No evidence for increased energy imports caused by Chinese ETS pilots [262]	Within firm reallocation of production to other jurisdictions found for Californian cap-and-trade scheme [287]; No evidence for reallocation to exempted firms by EU ETS [240, 254]			No evidence for leakage caused by Chinese ETS pilots [261, 288]; Positive spill-over effects caused by Tokyo and Saitama ETS [289]
Salience	The salience of the policy influences the responsiveness of firms and individuals to the policy		Possibility to sell emission allowances reduces emissions less than need to purchase allowances (EU ETS) [290]	Possibility to sell emission allowances reduces emissions less than need to purchase allowances (EU ETS) [291]	British Columbia carbon tax induces larger emissions response than market induced price changes [176, 231]	British Columbia carbon tax induces larger emissions response than market induced price changes [177]	Cross-country study finds introduction effect regardless of the carbon price level [18]
Anticipation	Emissions reductions depend on anticipated emission costs		RGGI caused emissions reductions already before its implementation [229]	EU ETS caused emissions reductions already before its implementation [240]	Downward adjustments of future carbon taxes lowered emissions reduction efforts in British Columbia [232, 292]		Chinese ETS pilots not found to cause emissions reductions between announcement and implementation [269, 288]

4.4.1 Hypothesis: Emissions are reduced by switching from high emitting fuels to lower emitting fuels

A common hypothesis, expressed in many of the reviewed articles, is that individuals or firms would respond to a carbon price by switching fuels [174, 178, 230, 239, 252, 255, 256, 266, 277, 279, 293, 294]. The amount of greenhouse gases emitted by the combustion of fossil fuels varies by fuel. In particular, coal is significantly more emission intensive than natural gas [178, 230, 252], oil can be placed somewhere between these two fuels [230], considering their global warming potential over 100 years. Therefore, carbon prices may induce firms to switch from fossil fuels with a higher emissions intensity to fuels with lower emissions intensities. (Switching from fossil fuels to low-emission alternatives is captured in hypothesis 4.4.2 below.)

Different sectors may have varying potentials for fuel switching, depending on the type of fuel used and the availability of alternatives in the short- and in the long-run. Particularly the power sector is commonly characterised by capacities that exceed the power demand most of the time. If these capacities are made-up of power plants using different fuels, a switch in utilised fuels is feasible in the short-run [178, 252, 293–295]. In the transport [252, 296] and buildings [177] sectors, on the other hand, users are commonly bound to a specific fuel for the lifetime of their vehicles and heating installations, which may only be substituted over a longer period of time. For the manufacturing sector, there are conflicting expectations as to whether fuel switching is feasible in the short-run [277, 293], likely related to the variety of technologies used in the sector.

Evidence: Evidence for fuel switching is only found for the energy sector in the UK, subject to the carbon price support [175, 178, 252, 253], replacing coal with natural gas combustion. The EU ETS and RGGI, instead of replacing coal by gas, are found to reduce coal combustion in the energy sector without an uptake in the combustion of other fuels [178, 230, 256, 257]. Consistent evidence for the UK carbon price support shows that the policy caused a significant decrease in emissions from coal power generation, while emissions from natural gas power plants increased at a lower magnitude [175, 178, 252, 253]. In fact, a few years after the introduction of the carbon price support the share of coal in the energy production was lowered from

39% (2012) to 2.4% (2019), with the carbon price ranging between US$ 30 and 40 in 2019 [252]. Evidence for the RGGI [230, 256, 257] consistently shows that the carbon price reduced emissions from coal power generation, but did not increase emissions from natural gas power generation. The same effect is found for the EU ETS by a study on the German energy sector [178].

Fewer studies are available for the industry sector, while the transport and buildings sectors are only covered by descriptive evidence. The industry sectors covered by the ETSs in the EU and Japan show evidence for relative fuel switching. The EU ETS is found to cause significant reductions in the emissions from coal combustion during the second phase of the ETS policy, while no switch to natural gas or oil is detected [240, 254, 255]. Yajima *et al.* [258] find no evidence for fuel switching caused by the Tokyo or Saitama ETS in the industry sector. Descriptive evidence from Sweden and Finland points into the direction that the carbon taxes also in the buildings and transport sectors led to fuel switching. In Sweden, the combustion of oil heating in households declined, while heating with wood, electricity or district heating increased [235]. A similar trend is observed in the transport sector, where the share of diesel cars compared to gasoline cars in Sweden increased substantially between 1991 and 2015 [259], while in Finland the consumption of diesel increased and the gasoline consumption dropped [260]. While the Swedish carbon tax was introduced with a price level of US$ 30, for Finland, this trend was observed starting at carbon prices below US$ 10.

The evidence for the Chinese ETS pilots is not sector specific, but provides evidence for relative rather than absolute fuel switching across sectors. The policy is found to reduce coal combustion, both in absolute terms [261, 262] and as a relative share in the energy mix [263–267]. There is less evidence available to test whether coal combustion was replaced by the combustion of other fuels. An increase in the relative share of natural gas combustion is detected [268, 269], but no absolute uptake of natural gas combustion [261]. When including also other energy sources (hydro, nuclear, wind and natural gas), the ETSs are found to cause an increased energy generation from these sources [262]. The causality between the ETS pilots and reduced coal combustion should be treated with some caution, as other Chinese regions, with high GDP per capita, are found to reduce coal combustion around the same time as the ETS pilots were introduced [261].

Some more detailed evidence from the energy sector, provides further insights into the preconditions for the fuel switching mechanism. Gugler *et al.* [178] calculate the marginal costs of each power plant at different carbon price levels to examine the "power supply structure (called the 'merit order')". They show that a carbon price of 25 EUR reverses the supply structure, such that enough gas power plants supply electricity at lower marginal costs than coal power plants to satisfy the entire electricity demand at commonly observed demand levels. A similar effect is not found for the German energy sector, where the carbon price of the EU ETS was too low (until 2018) to cause substantial fuel switching between coal and natural gas. For the RGGI states, fuel switching is only observed in periods of low natural gas prices, indicating that the carbon price is unable to reverse the merit order in periods of high natural gas prices [257]. The merit order is of lower importance in highly regulated or monopolised energy markets. Cao *et al.* [262] points out that electricity generation in China is based on quotas allocated without consideration of the generation costs, questioning the viability of fuel switching between facilities. And Leslie [295] shows, based on evidence for the Australian ETS, that large power producers, which run both coal and gas power plants and posses sufficient market power, may have an incentive to continue running coal power plants instead of switching to natural gas power generation and thus obstruct the market mechanism from causing a fuel switch.

Conclusions: Emissions are reduced by switching from high emitting fuels to lower emitting fuels, under the condition that lower emission technologies are available and able to substitute previous processes. A further condition is that the carbon price is high enough to reverse the marginal costs of different fuels in favour of the lower emitting fuel. These conditions are mainly met in the energy sector, while other sectors show little evidence for fuel switching in the short run. Relatively, carbon pricing is found to reduce the share of high emitting fuels across sectors and schemes. The reviewed studies largely focus on the first few years after a policy is introduced. There are some indications that over longer time periods some durable technologies are also replaced in favour of technologies using lower emitting fuels. The evidence from the UK suggests that substantial emissions reductions can be achieved via fuel switching already at moderate carbon prices.

4.4.2 Hypothesis: Low-carbon technologies are adopted to reduce greenhouse gas emissions

Carbon pricing may not only motivate the switch from one fossil fuel to another, but also incentivise the adoption of other energy sources, such as wind, solar, hydro or nuclear, or in demand-sectors the electrification of key processes [252, 266, 272, 279]. This mechanism is assumed to be a gradual transition [266], as the installed low-carbon technologies in the power sector are expected to run at (close to) full capacities even without the carbon pricing policy [252]. The mechanism, therefore requires the installation of new capacities.

Evidence: Evidence for this mechanism is limited to the energy and buildings sectors. Yajima *et al.* [258] find no significant effect of the Tokyo ETS on the energy produced from renewable sources. RGGI has not caused an increase in the energy produced by renewable energies, in fact the uptake of renewable energies in the RGGI states was significantly slower as compared to other US states that did not apply a carbon price [257]. When the UK carbon price support was introduced, the hydro power has partly filled the gap of reduced fossil energy production (roughly 15%) [252]. For nuclear power, Gugler *et al.* [252] find no significant uptake related to the UK carbon price support. For the buildings sector, Runst and Thonipara [235] provide descriptive statistics that around the same time when carbon taxes were increased, the sale of heat pumps more than doubled, without assessing the causal impact of the tax increase on the uptake of the technology.

Conclusions: The reviewed studies provide limited evidence of carbon pricing reducing emissions via the adoption of low carbon technologies. This might be related to the type of evidence reviewed. Most studies in our sample assess the effectiveness of the policy across a rather short period of time. As expressed in the hypothesis, this mechanism is likely to take some time to unfold. Further evidence to test this hypothesis can be expected to become available the longer the policies are in place. Another explanation could be the rather low carbon prices observed for many schemes in the primary studies, reducing the profitability of the adoption of new technologies.

4.4.3 Hypothesis: Emissions are reduced via efficiency improvements

Many study authors express hypotheses that carbon pricing may reduce emissions by lowering the emissions per unit of output [230, 240, 255, 264, 271, 274, 281, 293]. This may be achieved via reductions in energy intensity [264, 274, 293], reduced emission intensity of energy [269], or shifts in the economy towards industries with lower emission intensity [263]. Energy intensity could, for example, be achieved by "switching to a combined heat and power system" [230] or more generally by making "more efficient use of process heat" [240]. Reductions in the emission intensity of energy are closely linked to the mechanisms discussed above to switch to lower emitting input fuels or to replace fossil fuels by low-emission technologies such as renewable energies.

In the reviewed studies the emission intensity mechanism is described for the industry [240, 293] and energy [230] sectors. It is not excluded that the same mechanism could also be relevant in other sectors. As the reduction of emission intensity is expected to require investments [271], this mechanism is expected to require some time to unfold [230]. As the costs may differ by sector or firm, improvements in the emission intensity are expected to happen where the abatement costs are lowest [293], while in other sectors the abatement costs may be too high to observe any reductions in the emission intensity [283]. Also the size of the emitting facilities may have an impact on whether the emission intensity is reduced, as larger firms may have the advantage of scale effects so that an improvement in the processes may lead to larger emission savings and larger avoided payments for the carbon price [291].

Evidence: Evidence suggests that the ETSs in the EU, China, and Quebec caused significant reductions in the emission intensity of output. For the EU ETS this is particularly studied for the industry sector, while the policies in China and Quebec are studied across sectors. For the EU ETS, consistent evidence for reduced emission intensity is found for the manufacturing sectors in France [254, 255], Germany [240], and Lithuania [270]. Evidence for all three countries shows significant reductions during the second phase of the ETS policy. Only in the Norwegian manufacturing sector, the EU ETS is not found to cause significant improvements of the emission intensity [271], while there is a lack of evidence for other EU countries. Studies on the Chinese ETS pilots, conducted across sectors, largely agree that the policy achieved significant reductions in emission intensity [265, 267, 269, 272–274], while a few studies do not find significant improvements of the emission intensity [275, 276]. Qi *et al.* [265] study the emission intensity effect per ETS pilot and find significant emission intensity reductions only for Beijing and Hubei, while they do not find significant effects for Tianjin, Shanghai or Guangdong. The Quebec ETS in Canada is also found to significantly decrease the emission intensity [277]. Only for the RGGI, the evidence suggests an increase in the emission intensity in the power sector, likely caused by a utilisation rate below the capacity of each plant [257].

There is less evidence available to test whether this reduced emission intensity of output is explained by reduced energy intensity, reduced emission intensity of energy, or both. There are four studies testing the effect of the Chinese ETS pilots on energy intensity. Three find evidence for reduced energy intensity [174, 272, 273] and one finds an insignificant reduction effect [269]. The findings still show a limited consistency, with estimated reduction effects ranging from 1% to 39%. Cui *et al.* [269] furthermore finds a significant reduction in carbon intensity of energy caused by the ETS pilots.

Firm level evidence reveals that carbon intensive plants are more responsive to the policy and that firms react by improving efficiency or shutting down of their facilities. Kim and Bae [293] find a differentiated response to the ETS implementation by Korean manufacturing firms. They use the fact that almost half of the manufacturing firms before the policy implementation were certified for their environmental performance, while the other half did not hold such a certificate. The ETS implementation caused significantly higher emission reductions among firms without environmental certification compared to certified firms. This suggests that firms with a worse environmental performance before the policy implementation follow the best practices of those firms that had reduced emissions before the policy was in place. In the EU, firms are found to increase their investments into technology to reduce emission intensity [255] (further discussed in section 4.4.4). The carbon price in the UK is found to cause the shut down of the most emission intensive plants first, while later increases of the carbon price had relatively smaller reduction effects, as only more efficient power plants remained on the market [252]. The findings by Leroutier [175] further suggest that the UK carbon price support inhibited the entrance of high emitting plants to the UK energy sector.

Conclusions: The reviewed evidence supports the hypothesis that carbon prices achieve emissions reductions via efficiency improvements. The evidence is most comprehensive for reductions in the emission intensity of output, while only providing suggestive evidence that both energy intensity and emission intensity of energy are reduced. Firm level evidence suggests that most significant reactions to the policy are observed for plants with high emission intensities, which are driven to improve their efficiency or pushed out of the market. This mechanism could be studied in further detail and potentially be split into more fine-grained mechanisms to understand what actions firms are taking to reduce emission intensities and how these are motivated by the carbon price. Finally, we did not identify any evidence for this mechanism in the buildings and transport sectors.

4.4.4 Hypothesis: Carbon pricing fosters investments

The above presented mechanisms often require investments into updating or purchasing of equipment [271, 277, 278, 297], such as machines [271], heating systems [235] or vehicles [232]. Carbon prices may be taken into account for investment decisions as they have a direct influence on the profitability of different technologies over their lifetime. Investment decisions further depend on the availability and knowledge about affordable technical solutions [271] as well as the financing conditions for the investor [287, 297]. The mechanism is expected to reduce emissions in the long- rather than the short-run, as the translation of investment decisions into emissions reductions takes time [235, 271].

Evidence: In general, the evidence for this mechanism in the reviewed studies is quite scarce. Colmer *et al.* [255] study the investments of French manufacturing firms subject to the EU ETS and find a significant increase in investments caused by the carbon pricing policy. In particular, investments in less emission intensive processes and technologies are found to increase, while there is no significant effect detected for investments on emissions detection or end-of-pipe emissions reductions. They further support their findings from the investment data by conducting a survey of French manufacturing firms. Firms which are participating in the EU ETS are found to be more likely to invest into improvements in the use of process heat as well as process optimisations.

Only one study assesses the interaction between the policy and foreign direct investment. Their findings imply that, while foreign direct investment in China generally has an increasing effect on greenhouse gas emissions, this effect is reversed by the introduction of the pilot ETS policy [278]. Without the backing of further evidence this should however be treated with caution.

Some more descriptive evidence is provided by Runst and Thonipara [235] for the Swedish buildings sector. They find that heat pump sales substantially increase following a significant increase of the Swedish carbon tax rate from about US$ 50 to above US$ 120. During the same years, the residential per capita emissions decrease significantly. The authors point to an "adjustment lag", which supports the assumption that emission reductions achieved via investments will take some time.

Conclusion: The scarce evidence described by the reviewed studies provides a weak indication that carbon pricing fosters investments. One study from Sweden points to the critical role of the carbon price level to incentivise investments. Again, the limitation in the available evidence is likely explained by this mechanism requiring time to unfold, while the reviewed studies are largely focused on the first few years after the policies were introduced. Further evidence would need to be collected to test this hypothesis.

4.4.5 Hypothesis: Carbon pricing spurs research and development to achieve emissions reductions

Besides investment in available technologies, the policy is expected to incentivise an uptake of research and development to find new solutions for emissions reductions [230, 261, 263, 264, 274, 275, 277, 279, 298, 299]. The success of these activities to develop scalable solutions for carbon intensive processes is not guaranteed and is expected to take time before resulting in emissions reductions [261].

Evidence: This mechanism is only studied for the Chinese ETS pilots among the studies we review. The evidence shows weak indications for increased expenditure for research and development as well as for an increased number of issued patents. Early evidence also suggests that increased research activities have played a role in bringing emissions down.

Out of three studies assessing the expenditure, two find an increase [264, 279] (the latter assesses an index combining expenditure with further research and development indicators), while one study does not find a significant effect of the ETS pilots on the expenditure for research and development [261]. Yang *et al.* [298] finds that also the governmental expenditure for research and development significantly increased with the implementation of the policy. It is, however, difficult to argue that governmental spending can be the result of a policy.

With respect to issued patents, Xu [263] find that the Chinese ETS pilots have a positive effect on the total number of issued patents. Dong *et al.* [268] divide the patent data into "green inventions" and other inventions (without a transparent definition for green inventions) and find that the ETS pilots specifically promoted "green" innovation. This is partially supported by Chen *et al.* [299], who find mixed results for different indices of low-carbon technology patents.

Xu [263] and Wang *et al.* [264] further test, whether achieved emission reductions are caused by the increased innovation efforts using mediation analysis. Both studies find that the uptake in innovation is a causal mechanism towards emissions reductions. The studies assess data for the first seven and three years after the policy implementation, respectively. This either implies a very rapid success of the innovation mechanism or questions the reliability of the findings.

Conclusion: The reviewed evidence provides a weak signal of an uptake in research and development activities fostered by the Chinese ETS pilots. Only some of the studies find statistically significant effects for some measures of the research and development activities. Early assessments of the causality of this mechanism for emissions reductions imply a causal chain from carbon pricing via research and development leading to emissions reductions. The significance of this mechanism should be further substantiated with evidence from other carbon pricing schemes and studies of longer time periods.

4.4.6 Hypothesis: Emissions are reduced by reducing production

Instead of adjusting the production processes, firms and households could also avoid the costs induced by carbon pricing by scaling down their production or consumption [230, 257, 264, 271, 275, 277, 279, 281, 296]. The carbon price induced cost may make the production of specific products unprofitable [271] or deteriorate their competitiveness with respect to products that are not subject to the policy [298], causing firms to reduce their production. If the costs are passed on to consumers, they may also adjust their consumption accordingly [296].

This mechanism is mainly expected in the manufacturing [271, 277] and energy [230, 257] sectors. As the demand for heating and transport are assumed to be inelastic, this mechanism is not assumed to significantly affect the buildings and transport sectors [231]. At least for the aviation sector, Heiaas [296] challenges the inelasticity assumption and instead assumes reduced demand for air travel once ticket prices are increased.

Reduction in production is a flexible response to policy signals and may be temporary until a more permanent solution is achieved via one of the other mechanisms with a longer time lag [277, 296]. Carbon induced costs are most significant for emission intensive firms, which are therefore expected to be the first to reduce production, potentially being replaced by less carbon intensive industries [263]. A reduction of production by firms subject to the policy may be replaced by firms that are not regulated by the same policy [230]. This mechanism is discussed in section 4.4.7 below.

The assumption of reduced production may also be reversed. As discussed above, the policy may foster the adoption of more efficient technologies. These in turn may have a positive impact on production [263, 281].

Evidence: The reviewed evidence does not suggest significant reductions in overall production levels. Measured by the gross domestic product (GDP), no evidence for reduced production levels is found for the Chinese ETS pilots [265, 269, 281, 282, 284], British Columbia carbon tax [285], or carbon taxes applied across European countries [286]. Only one study challenges the overall consistent evidence for China, finding some indication for a short-term reduction in gross domestic product in the first year of the policy [300], while the significance of the finding (from a synthetic control design) is unclear.

Other studies instead proxy the overall output using data on employment levels. From significant reductions in employment caused by the Quebec ETS, Hanoteau and Talbot [277] conclude negative effects of the policy on production. The Chinese ETS pilots, on the other hand, are found to have a positive impact on employment levels [298]. Changes in employment levels are not necessarily an indication for changes in the level of output and should only be interpreted as suggestive evidence for a change in production.

Sectoral level estimates show that industrial output was not affected by carbon pricing, while other sectors were triggered to reduce their production. Industrial output is measured in accordance with the gross domestic product examined above. It is not found to be significantly reduced by the Chinese ETS pilots [281, 282] or the EU ETS [240]. Sector specific measures of economic activity, show reduced energy generation caused by the Chinese ETS pilots [174, 262, 269, 273], but not by the Tokyo ETS [280]. Descriptive evidence for the UK carbon price support shows a modest decline in electricity consumption of about 5% during the first few years of the policy [252]. The EU ETS is further found to cause a significant reduction in short-haul flights below 1000km distance [283].

Heterogeneous responses in gasoline demand are found for the transport sector in British Columbia. While the carbon tax is found to reduce the demand in larger cities, more rural areas show no response to the carbon pricing policy [231]. This could be explained by more readily available alternative modes of mobility, such as public transport, cycling, or walking.

Conclusion: Carbon pricing is not found to generally reduce production, but specific sectors react to the policy by scaling down their activities. The evidence does not reveal why the sectoral declines are not showing up in the larger aggregates. This could be related to measurement uncertainty or the size of the effect being too small compared to the aggregate. More interestingly, it might also be explained by shifts in economic activities, which are insufficiently studied. Declines in flights or car use may, for instance, be substituted by an increase in train travel. This should be further investigated by future research for a better understanding of the mechanism.

4.4.7 Hypothesis: Carbon prices cause 'leakage'

Firms can also try to avoid policies by relocating their production and associated emissions to other jurisdictions or facilities not subject to the policy [178, 230, 233, 255–257, 261, 262, 277, 287–289, 298, 301, 302]. This phenomenon, known as leakage may be observed in different forms. Firms operating both within and outside the carbon price jurisdiction could relocate (parts of) their production to facilities outside the regulated area [255, 287, 289]. Leakage can also occur when actors within the carbon pricing jurisdiction take the decision to reduce production, resulting in an increased import from or decreased exports to non-affected jurisdictions [240, 288, 289]. Even within the jurisdiction the policy may incentivise firms to reallocate their production, if some actors (e.g. based of the size of the facility) are exempted from the carbon price [254, 255].

The leakage mechanism is available in the short-run if spare production capacities [178, 287] and, in the case of the energy sector, transmission capacities [178] are available. It might be a temporary response to carbon pricing until other emission reduction mechanisms, which take longer to implement, become operational [230].

Some sectors are more prone to emission leakage than others [302]. In particular, the energy [256, 257] and industrial [288] production can be relocated. In the transport sector, leakage is limited to cross-border fuel shopping [231] and in the buildings sector the leakage mechanism is likely unavailable.

Leakage can offset achieved emissions reductions and potentially increase global emissions [255]. Counter to the leakage hypothesis, carbon pricing could also cause emissions reductions outside the treated jurisdiction, if enhanced procedures and technologies are adapted by those not subject to the policy or if emissions reductions outside the regulated area can be accounted for to lower carbon price expenses (offsets) [233, 261, 289].

Evidence: Evidence on emission leakage is found for the Californian cap-and-trade scheme [287], for the RGGI [230, 256, 257], and for the UK carbon price support [252]. Studies on the EU ETS [240, 254, 255], Chinese ETS pilots [261, 262], Tokyo ETS, and Saitama ETS [289] do not show evidence for emission leakage. Firms that operate in California and also have plants outside of California are found to shift production from Californian plants to facilities not subject to the cap-and-trade policy [287]. For both the UK and the RGGI participating states, evidence shows that some of the reductions in electricity generation from coal and natural gas within the policy area are replaced by energy generation from outside the policy area [230, 252, 256, 257].

For the EU ETS, different lines of evidence are examined, with none pointing towards leakage. Wagner *et al.* [254] and Colmer *et al.* [255] examine emission leakage within the EU. As the policy entails exceptions, particularly, based on the size of the facility, emissions could be reallocated from larger to smaller facilities. Both studies find no evidence for within EU leakage in the French manufacturing sector. Colmer *et al.* [255] furthermore finds no evidence for emission leakage via increased imports of emission-intensive intermediate products from outside the EU. French manufacturing firms are not found to increase their import volumes. Petrick and Wagner [240] examine whether emissions of German export goods may be relocated by production outside the EU. The study finds no reduction in German manufacturing exports, but cautions the interpretation of the findings, as the studied data include export data to other EU countries, which are subject to the same policy. Chinese ETS pilots are not found to cause significant increases in electricity imports or increased power generation in neighbouring countries [262]. Also for other sectors, Yang *et al.* [261] do not find any effects on increasing emissions in areas adjacent to the pilot regions. Gao *et al.* [288] calculate production based and consumption based emissions for the pilot regions. The study finds slightly larger emissions reductions for production based emissions, but does not test whether the difference is statistically significant.

The total effectiveness of the policy beyond the regulated facilities depends on the relation of leaked emissions to the achieved emissions reductions. Evidence for California shows that the policy caused emissions increases outside the cap-and-trade area, which outweigh the achieved emissions reductions within California [287]. In the UK and RGGI states, the imported electricity had a lower emission intensity than the replaced (mostly coal) electricity. For Japan, the evidence even suggests a positive spill-over effect from the ETS policies in Tokyo and Saitama. Firms subject to one of the ETS policies are found to reduce their emissions not only in the facilities within the regulated area, but also in facilities that are not subject to the policy [289].

Conclusion: The emission leakage mechanism is found to play a role for some, but not all, carbon pricing schemes. Evidence for leakage is found for the UK and RGGI energy sectors and the Californian industry sector. The reviewed studies, however, do not allow to identify which context factors would support or hinder leakage. In particular, the case of the Chinese ETS pilots shows that proximity and economic integration of regulated and non-regulated areas alone do not suffice as explanations. It should be further studied which other context factors are relevant to trigger the leakage mechanism and how policy designs may avoid negative consequences of leakage. Given that the studied leakage effects take place in a context of comparatively low carbon price levels, further evaluation is needed to understand to what extent higher carbon prices intensify the leakage mechanism, undermining global emissions reductions.

4.4.8 Hypothesis: The effectiveness of carbon prices is related to their salience

The above mechanisms require firms and individuals to take action. It is hypothesised that the likelihood of these being triggered is correlated with the salience of the carbon price [176, 232, 303]. Prices for fossil fuels are commonly quite volatile, such that the effectiveness of a carbon price is argued to depend on the degree with which this price change is recognised by market participants [176, 232].

Evidence: Three lines of evidence can be interpreted to support this hypothesis. First, the reaction to a carbon price significantly exceeds the response to other fuel price changes. Evidence from British Columbia shows that for both the transport [176, 231] and the buildings sector [177] the semi-elasticity of fuel demand is significantly higher for carbon price changes than for market price changes.

Second, the introduction of a carbon price is found to cause an introduction effect unrelated to the size of the carbon price. This is based on a cross-country evaluation of carbon prices introduced at various price levels [18].

Lastly, [291] and [290] show that firms with abundant emission allowances reduce their emissions less than firms with a shortage of emissions allowances, even though the opportunity cost of an emitted ton of carbon dioxide is the same for the purchase or sale of an allowance. This evidence shows that the higher salience of the carbon price for the purchase of an allowance causes larger emissions reductions than the forgone opportunity to sell an allowance at the same price.

Conclusion: The evidence provides some indication that the effectiveness of carbon pricing is related to their salience. Further research is needed to better understand the relevance of this mechanism and potential implications for the communication of the policy.

4.4.9 Hypothesis: Emission reductions depend on anticipated carbon prices

Any reaction to a carbon pricing policy is assumed to be dependent on the expectations of future carbon prices [232, 277, 302]. This is particularly the case for mechanisms that require long-term planning, such as investment or research and development decisions [277]. If a carbon price generates expectations of permanently increasing fuel prices, this will be accounted for in the decision making [232]. This mechanism may, however, be reversed if firms expect to have an influence on the future stringency of the policy, such that lower abatement efforts today would lead to looser emission caps in future.

Evidence: There are two lines of evidence providing an indication of the relevance of expectations for emissions reductions. Murray and Maniloff [229] and Petrick and Wagner [240] find evidence for an emissions reduction effect already before the policy was implemented, while Pretis [292] and Erutku and Hildebrand [232] find lowered emission reduction efforts when expected tax increases did not materialise.

For both the EU ETS and the RGGI, the policy was announced respectively two and three years before coming into force. This resulted in significant emissions reductions in the affected sectors already before the carbon price was charged [229, 240]. Gao *et al.* [288] and Cui *et al.* [269] test for the same effect for the Chinese ETS pilots and do not find significant emissions reductions during the two years between the announcement and the implementation of the policy.

Revoked carbon tax increases were found to decrease the achieved emissions reductions in British Columbia. The policy was scheduled to regularly increase the tax rate by a fixed amount each year. This plan was revoked after the first four price raises and the tax was held constant between 2012 and 2018. Pretis [292] and Erutku and Hildebrand [232] find that the non-occurrence of the tax increase in 2012 coincided with a decreasing effectiveness of the policy and a rebound of emissions in the transport sector.

Conclusion: The limited evidence evaluated here, supports the hypothesis that carbon pricing effectiveness is related to the anticipation of future price levels. Further research should be assessed to better understand the carbon price expectations of different actors and how they interact with the mechanisms for emissions reductions.

4.5 Discussion

Here we provide the first realist synthesis for rigorously understanding climate change mitigation policies. With an application to carbon pricing policies, we showcase the relevance of systematic assessments to understand what climate policies work, under what conditions, and why. Carbon pricing is commonly applied across sectors and praised for its flexibility to incentivise emissions reductions where and by what means they are the easiest to achieve [165]. The multiple potential mechanisms at play spark heated debates on whether and how carbon pricing is able to reduce emissions. In this realist synthesis we systematically track hypotheses on how the policy works and comprehensively evaluate to what extent these are supported by the available evidence. We synthesise 177 hypotheses from 80 carbon pricing policy evaluations and identify nine main mechanisms how carbon pricing works and a range of context factors, which make it more or less effective. We assess 293 pieces of empirical evidence to test the validity of each of the hypotheses across the contextual variations from 21 policy schemes and four emissions sectors. We provide a refined theory of the causal chains between the application of carbon prices and the reduction of greenhouse gas emissions.

Meta-analyses have established that carbon pricing effectively reduces greenhouse gas emissions, with varying and largely unexplained variations between schemes (see Chapter 3 and [304]). How these emissions reductions come about has previously not been systematically assessed. While most of the mechanisms assessed here have been the subject of previous reviews, none has applied a systematic method for the selection and synthesis of the reviewed studies. Previous reviews on carbon pricing have covered fuel switching [305], low-carbon technology adoption [138–140], investments [139, 305, 306], research and development [139, 140, 305, 306], reduced production [138, 307], and leakage [308]), but do not attempt to systematically study how these mechanisms unfold in varying country and sector contexts.

In Chapter 3, we show that carbon prices are generally effective in reducing greenhouse gas emissions, but achieved emissions reductions vary substantially. This provided the motivation to study the mechanisms that underlie emissions reductions and how these vary across heterogenous contexts. The evidence reviewed here suggests that emissions reductions are in the short-term achieved by a mix of fuel switching, efficiency improvements, downscaling of some emission intensive activities, and leakage. The prevalence of each of these mechanisms is highly context dependent and we identify a number of context factors preventing the mechanisms to unfold. For other mechanisms the evidence is considerably thinner. The reviewed evidence is particularly scarce to study to what extent carbon pricing has caused a transition to low-carbon technologies, foster investments, or boost research and development activities. This may partially be explained by the focus of the reviewed studies on the immediate policy responses.

We also limit ourselves in this review to studying the mechanisms in isolation, while in reality they should be assumed to interact extensively. These interactions need to be studied further. We are also still at the beginning of identifying which context factors influence the occurrence of each mechanism.

From the reviewed evidence, we gain limited insights on more gradual mechanisms and the interaction of carbon pricing with other policies and external events. Quasi-experimental study designs, contributing the largest part of the evidence reviewed here, are well suited to study abrupt responses to an intervention, but provide fewer insights in more gradual changes achieved over longer time periods. This likely explains the focus of the reviewed evidence on more immediate mechanisms as compared to mechanisms requiring more fundamental changes, e.g. in the development and adoption of new technologies. These need to be studied further using other lines of longer-term evidence. Further research is also required to study the relevance of other policies or external shocks that may support or obstruct the effect of carbon pricing. In a number of cases the introduction of a carbon pricing scheme, for example, coincided with an economic recession (EU ETS, RGGI, and carbon taxes in Sweden, Finland, and British Columbia) [176, 229, 240, 259, 260, 271] or natural disaster (Tokyo and Saitama ETS) [301]. The reviewed studies provide little evidence how these events have interfered with the carbon pricing mechanisms. Similarly, there is a growing interest in the synergies and conflict between policy instruments [9, 118, 309, 310], but our sample provides little insights in the interaction of carbon pricing with other policies. The theory generated here needs to be further refined taking into account the longer-term impact of carbon pricing as well as its interaction with parallel events and policies.

Many of the evidence gaps identified here, could likely be filled by expanding the amount of considered evidence and by more comprehensively integrating heterogenous lines of evidence from qualitative studies, surveys, descriptive studies, etc. The 80 studies reviewed here, provide a rich set of evidence on a range of context-mechanism-outcome configurations, while others are not (sufficiently) studied. The realist synthesis approach encourages to iteratively search and review evidence, identify new hypotheses arising during the synthesis, and test these by reviewing further lines of evidence. Considering the massive amount of potentially relevant evidence, indicated by the more than 4,000 carbon pricing policy evaluations identified in Chapter 2, this may easily grow into an infinite review project. Realist synthesis therefore challenges the reviewer to find the right balance between the comprehensiveness of the review and the feasibility of the project. Given the many unanswered hypotheses that could be derived from the synthesis provided here, we find it difficult to conclude the project at this stage, but are positive about the contribution this first realist synthesis can bring to enhance our understanding of carbon pricing. The framework of contexts and mechanisms developed here may be used to further refine the evidence-based theory by integrating further lines of evidence.

The realist synthesis could also be extended to study the relevance of the carbon price level on each of the mechanisms. While it should be assumed that a higher carbon price leads to increased emissions reduction efforts, it may influence different mechanisms at different degrees. The studies reviewed here do not provide sufficient evidence to study the effect of the carbon price level in detail. The theory developed here could be linked with the concept of abatement costs [311–313] to study how much abatement could be achieved via which mechanism at a given carbon price level and context. It should however be noted that the majority of studies in our sample assess carbon pricing schemes with average prices below US$ 30 during the study period. With an increasing number of carbon pricing schemes with significantly higher prices, the evidence base should be strengthened to assess how the mechanisms unfold with higher carbon prices.

With this realist synthesis we aim to pave the way for more configurative systematic reviews in the field of climate policy evaluation. Yet, in this first realist synthesis in the field we have not met all criteria for a gold-standard systematic review. In particular, we have not developed a detailed protocol in advance of the review and have not conducted a rigorous critical appraisal. While it is good practice for any systematic review to develop and publish a protocol outlining all steps of the review process to ensure a transparent and objective assessment of the evidence [23, 24], we had to develop the review method applied here while conducting the review. The adaptation of realist synthesis from other disciplines, left us with limited examples of how this can be applied to the evaluation of climate policies, preventing us from developing a sophisticated protocol early on. Similarly, we had no experience or examples how critical appraisal can best be applied in a realist synthesis of climate policy evaluations. The appraisal in realist synthesis should combine the assessment of the relevance and rigour of the reviewed evidence, but there are no common tools available, how this should be conducted [314]. We therefore evaluated the reliability of the evidence without a pre-defined appraisal tool and report caveats as required.

Realist synthesis is a promising tool to facilitate evidence-based policymaking. It provides nuanced insights to understand how and under what conditions a climate policy measure works, forming a sound knowledge base to inform decision makers. Here, we provide a starting point for understanding which mechanisms are triggered by carbon pricing policies in different contexts. This should further be refined, integrating further lines of relevant evidence to derive a robust theory of how the policy works. The assessment framework developed here can be adapted to assess other climate policies and be extended to study further policy outcomes. Policymakers require a sound knowledge of how climate policies can achieve substantial emissions reductions and how they affect other policy aims. We need to make the best use of the available evidence to understand what climate policies work, under what conditions, and why, to stand a chance of achieving emissions reductions at a pace required to stay within the 1.5°C or 2°C warming limits [1].

Chapter 5

Discussion

In this thesis, I adopt three evidence synthesis methods and apply these to study what the literature knows about carbon pricing policies. Here, I want to reflect what can be learned from the three applications to answer the research questions set out in the introduction (see section 1.4). The question what benefits evidence synthesis methods bring to climate policy evaluation is discussed in section 5.1, followed by a stocktake of remaining challenges for evidence synthesis to be more widely applied for climate policy evaluation in section 5.2. Finally, in section 5.3, I reflect on how evidence synthesis can benefit the research community, policymakers, and global environmental assessments.

5.1 Benefits of evidence synthesis methods to evaluate climate policies

One of the main motivations for this thesis was to assess, how well evidence synthesis methods can be applied in the field of climate policy evaluations and what benefits their application entails. Based on the three methods used for this thesis, I aim to assess their value for researchers, policymakers, and environmental assessment bodies. Remaining challenges are discussed in section 5.2. The discussion of the benefits is structured by the three evidence synthesis methods applied in Chapters 2-4.

5.1.1 Mapping provides an accessible repository to guide research and policymaking

Evidence and gap mapping was invented in the field of international development as a tool to make the best available evidence accessible to policymakers and to organise research agendas around existing gaps in evidence [82, 86]. For policy-relevant research findings to be picked up in policy-making processes, decision-makers need to be guided to the best available studies on an upcoming policy question [87, 315–317]. If, instead, these studies are hidden in large piles of irrelevant studies and search engines provide no purpose specific filters, valuable research findings are at high risk to be overlooked in policy-making [86]. Equally, for researchers, research commissioners, and science assessment bodies a structured overview of a research field can be helpful to scope the literature, ensuring that the research question, set out for a commencing project, is filling a research gap or retrieving relevant research to be synthesised [82, 86]. This reduces inefficiencies from duplicated efforts in searching and screening the same literature across project teams and allows for a transparent assessment of the research priorities [86]. Chapter 2 builds on the mapping experience in international development and adopts the method to the field of climate policy evaluations.

Policy-relevant research should be structured along the lines of the policy instruments and outcomes to provide policymakers a comprehensive overview of the policy options at hand [9, 26, 86, 140, 318]. In Chapter 2, we therefore develop a typology of policy outcomes, synthesising the typologies of the IPCC [9] with other outcome typologies [140, 150–153], capturing 14 categories of policy outcomes. Additionally we capture the policy scheme, containing information on the implementing jurisdiction and the policy design, as well as the emission sector and method used for policy evaluation.

In comparison with previous maps of climate change literature, this results in a higher level of detail captured for each study, but a narrower focus on ex-post evaluations of only one policy instrument. Besides a very large map of the entire climate change literature clustered by machine learning [14], there are two climate policy maps available for mitigation [15] and adaptation [109]. Both policy maps cluster the evidence by categories of policy measures and provide a broad methodological distinction between ex-post and ex-ante studies. We here apply a more fine-grained categorisation of methods and policy schemes and add the crucial dimension of labelling policy outcomes. By only listing ex-post carbon pricing studies, we reduce the amount of captured literature, in comparison with the map of all climate change mitigation policies [14] to about 5%.

To provide the best usefulness for policymakers and researchers, the fine-grained clustering, applied in Chapter 2, should be extended to the entire body of 80,000 climate policy studies, identified by Callaghan *et al.* [15]. The typologies of policy outcomes and methods, developed in Chapter 2, are meant to be transferable to research on other climate policies. Extensions would be required to capture ex-ante methods and better distinguish qualitative research methods. The typology on policy outcomes may require small adjustments or additional categories to be transferable to other policy instruments, but should generally be applicable.

Comparing the evidence and gap map compiled for this thesis, instead, with the common practice in international development research (for an overview of evidence and gap maps, see [131]), we extract less information from the listed studies and do not apply a critical appraisal. Evidence and gap maps commonly provide summaries and critical appraisal of the systematic reviews captured by the map [86, 131]. With the expected rise of systematic reviews in the field of carbon pricing policy evaluations, this would be a valuable extension of the provided map and may be added to the interactive online repository with the first update of the living map.

Another noteworthy deviation from the evidence and gap mapping practices in other research fields is the limitation of the data collection to the titles and abstracts of the articles, instead off accessing the full texts of the studies. This is not commonly advised in the mapping guidelines [26, 86, 130]. I here instead follow the practice of previous large-scale maps in climate research [14, 15, 109, 319, 320], who map studies based on the information in the titles and abstracts. This is explained by the constraint time resources and by the potential to automate this process. Not only that the time which would have been required to label more than 4,000 studies at full text, would have exceeded the time I had available for this thesis, but also the climate research community has constraint time to map out tens of thousands of studies. Secondly, previous mapping studies have argued that the automation of the mapping task is easier with the more condensed information provided in titles and abstracts [88, 90, 319, 320]. This supports our attempt to automatically update the map.

The decision to code data only from the titles and abstracts marks a relevant trade-off that needs to be considered in future mapping studies. It was previously discussed, whether evidence and gap maps should aim for the same comprehensiveness in searching for evidence as other synthesis methods, and suggested that a more “rapid” approach to searching was justified by the aim of the method to provide an overview of the available evidence in a reasonable amount of time [86]. I would build on this argument to propose that mapping out the evidence in a large and growing research area could allow to weigh the time required high enough to sacrifice on the detail of the information that may be captured. Considering that the evidence and gap map will often not be the only source of information for any user, may allow for a reduced accuracy. This, however, certainly comes with the drawback that some information from a number of studies may be missed [321, 322]. Here, we particularly noted that qualitative research methods are less clearly identifiable from the abstracts than quantitative methods, as qualitative studies often do not mention a method in the abstract.

Both researchers and policymakers should generally have access to the most recent evidence. Providing a static repository of evidence would soon be outdated and diminish its usefulness. The map compiled here is one of the first attempts in climate research to provide a living evidence base. Fully automated updates of the evidence and gap map would be able to identify at least three quarters of newly published relevant studies and add them to the map. The precision and recall for these predictions may be adjusted based on the use case. As a researcher, I would certainly prefer the machine to include a few hundred falsely added irrelevant studies instead of missing some relevant ones (lower precision, high recall). A policymaker instead may have limited time to deal with many irrelevant studies and would tend to accept missing a few more relevant studies (higher precision, lower recall).

The usefulness of the here compiled map for the use in systematic review projects is currently tested by a number of researchers. Currently, it is used to plan and conduct systematic reviews on the carbon intensity outcome and the cost effectiveness of the policy. The map informed the review teams that comprehensive evidence is available for both research topics. It is further used as an initial set of potentially relevant studies for screening and to inform the development of search queries for these projects. Also other review authors, who started their reviews by standard searches of bibliometric databases, use the map as an additional source of studies and to evaluate the scope of studies identified via the searches. This is currently used for systematic reviews on the innovation and leakage outcomes [323, 324].

A reasonable judgement about the value of the evidence and gap map for policy-makers, research funders, assessment bodies and future primary research can only be gained with time. So far, the map was not shared with policymakers, funders, or assessment bodies. Whether the literature overview is presented in a helpful way for all these user groups remains to be seen. In section 5.3.2, I sketch a vision how evidence maps can facilitate the work of assessment bodies like the IPCC. The map is also meant to guide future primary research towards closing research gaps. Whether researchers and research commissioners use the map to identify gaps is not yet evident.

5.1.2 Meta-analysis provides first rigorous assessment of carbon pricing effectiveness

The systematic review in Chapter 3 synthesises quantitative evidence on the effectiveness of carbon pricing. This mainly promotes research progress for global environmental assessments and the research community, while the benefits for policymaking are secondary. For the study we collect and review the evidence from 21 carbon pricing schemes to answer the question: Is this policy measure effective? And, if so, how large emissions reductions does it achieve? The review takes a very global perspective, providing limited guidance for policymakers whether and how the policy could also work in their local context.

In terms of research progress, I need to compare the systematic review with a long list of reviews conducted on the topic. A number of reviews studies the effectiveness of carbon pricing. Most of these assess the emission reduction effect among a longer list of policy outcomes [138–140, 241, 305, 306, 308, 325]. Haites [184], Rafaty *et al.* [18], and Green [137] have previously reviewed the carbon pricing studies with a specific focus on the emissions reduction effect. These reviews, however, do not use a clear methodology for synthesising the evidence and do not cover the relevant literature comprehensively. A more rigorous review of the effectiveness, comparable to the systematic review, conducted in Chapter 3, was published shortly after the publication of the review conducted here [304].

Reviews should be clear in what research question they aim to study and comprehensively search for all relevant literature to answer this research question. Studying several policy outcomes in the same review, drastically increases the required workload, if all aspects are to be rigorously reviewed. In previous literature reviews on multiple carbon pricing outcomes, the broader range of reviewed outcomes is instead traded-off against the comprehensiveness of the reviewed literature on each policy outcome. For some of the early reviews on specific carbon pricing schemes [306, 308] comprehensiveness might still be feasible, but more recent reviews, studying multiple outcomes of dozens of carbon pricing schemes, are certainly overburdened by the amount of available literature that should be relevant to answer their review questions across outcome dimensions [138, 140, 241, 325]. This is clearly shown by the large amounts of carbon pricing evaluations identified in Chapter 2.

None of the previous reviews uses systematic searches to identify the relevant literature and it is usually unclear, whether a study, not mentioned in the review, was not found or explicitly excluded by the authors. A number of reviews, at least, report on the search terms and bibliometric databases used for searching [137, 139, 140, 241, 305, 308]. These searches commonly use very restrictive search terms [140, 305] and are conducted via Google Scholar rather than bibliometric databases with advanced search functions [137, 139, 241, 305]. A comparison of the list of studies identified for the systematic review, conducted in Chapter 3, with the most comparable review (same research question, recently published) by Green [137] reveals that the cited review misses at least 30 studies published at the time of the review. The missed studies include 20 evaluations of the Chinese ETS pilots, introducing a considerable bias towards European and North American carbon pricing schemes. Other reviews do not explain their method for searching and selecting the literature covered in their review [18, 138, 184, 306, 325]. They likely follow no systematic approach for searching the literature and have no explicitly defined criteria to decide what primary study is or is not relevant for the review.

Previous reviews do not set out, how evidence is extracted, harmonised, and appraised. The reviews also do not consider whether publication bias skews the reviewed study findings. The way primary evidence is reported in traditional literature reviews usually depends on how it is presented in the primary studies rather than the reviewer defining what evidence is best suited to answer the review question. This results in a large focus on main findings and often does not allow to systematically compare findings across studies, if they are reported in different metrics. None of the previous reviews systematically appraises the primary evidence. A rigorous assessment of the study designs allows to differentiate how reliable a given research finding is. The best assessment of study designs, used for the evaluation of carbon pricing, is provided by Vrolijk and Sato [241], but this does not feed into the review of the effect sizes. Chapter 3 shows that 46% of the carbon pricing effectiveness studies have medium or high risk of bias in their study design. Publication bias is also shown to have a significant influence on the available evidence and we need to assume that insignificant study findings are systematically missing in published research. This can only be identified using meta-analytic methods, not attempted by previous reviews. Here, we are able to control for risks of bias in primary studies and biases arising from missing insignificant studies.

Literature reviews commonly conclude that carbon pricing effectively reduces emissions, without providing a clear estimate how large the emissions reductions are [139, 140, 184, 325]. Some reviews focussing on the EU ETS, provide rough ranges of the emissions reduction effect achieved by the scheme [305, 306, 308]. A quantitative summary of the effect sizes across policy schemes is only provided by Green [137], who summarises the estimated emissions effect as a range of annual reductions of between zero and two percent. As the review provides no synthesis methodology, I am unable to replicate these numbers. From the systematic review, performed in Chapter 3, I know that reduction effects are commonly reported as average treatment effects and a transformation of these effects into annual growth rates, as attempted by Green [137], is non-trivial. Green's numeric summary of the findings can therefore not be verified.

The IPCC in its most recent assessment report [9] builds on a number of the above reviewed literature reviews as well as a few additional primary studies and concludes that "[t]here is abundant evidence that carbon pricing policies reduce emissions." Based on three primary studies, it quantifies the emissions reduction effect for the EU ETS to be "at 3–25%" [9]. No other quantification of the effectiveness of any other scheme is provided, leaving the achieved reduction effects of more than 70 carbon pricing schemes unaddressed. The meta-analysis, provided in this thesis, would have allowed the IPCC for a more in-depth assessment of the heterogeneous emissions reduction effects achieved by carbon pricing schemes implemented around the world, adding evidence for 20 additional carbon pricing schemes and reducing the uncertainty of the estimate provided for the EU ETS to between -4.0 and -10.5% (95% confidence interval).

The systematic review performed here not only adds new knowledge in quantifying the effects across 21 carbon pricing schemes, it also provides a more robust account of the available evidence on carbon pricing induced emissions reductions than any previous review. The comprehensive search of the literature more than doubles the amount of considered primary studies as compared to the previously most comprehensive effectiveness review [137] and reduces the bias towards European and North American policies by uncovering 35 research articles on Chinese carbon pricing policies. By systematically extracting, harmonising, and appraising primary research findings, we provide a traceable account and rigorous assessment of the available evidence. Our transformation of effect sizes to percentage changes in emissions makes the estimates comparable across studies. We report considerable risk of bias and publication bias in the reviewed literature that needs to be accounted for in any assessment of the research findings. Our assessment of the heterogeneity in effect sizes does not provide any easy answers to explain the varying effectiveness across schemes, but remains as an open research question to be addressed by further research (including the realist synthesis performed in Chapter 4).

Shortly after the systematic review presented in Chapter 3 was published, a complementary systematic review on the effectiveness of carbon pricing policies was published by Ahmad *et al.* [304]. While the main finding of their review, that carbon pricing effectively reduces emissions, agrees with the findings presented in Chapter 3, they highlight a difference in the effectiveness of carbon taxes and cap-and-trade schemes, not found in the meta-analysis presented here. Ahmad *et al.* [304] identify 81 relevant studies, with significant but not complete overlap with the 80 studies collected in Chapter 3. Their finding of differences in the effectiveness of taxes and trading schemes results from a comparison of means and is not supported by our meta-regression results including control variables. As the authors do not publish their data and have not responded to my inquiry, I am unable to investigate the difference in our findings more comprehensively. The two reviews further differ in that Ahmad *et al.* [304] transform collected effect sizes into correlation coefficients, not allowing for an easy interpretation of their findings, and do not report scheme-wise effects. Their review nicely complements the meta-analysis conducted here. Disagreements of our findings should be investigated to further enhance our understanding of the effectiveness of carbon pricing.

The significant advantages of systematic reviews for research progress and IPCC assessments do not directly translate into comparable advantages for policymaking. The systematic review, conducted in Chapter 3, addresses a narrow research question, evaluated against the experience from across countries and contexts. For some policy questions a rigorous primary study on a given policy implementation in its specific country context may be more insightful than a global review on the same question. Primary studies in addition often provide detailed insights into country-specific mechanisms and context factors, examine scheme-specific policy design features, or suggest policy amendments based on supplementary evaluations. Our meta-analysis, instead, only shows the effectiveness of carbon pricing and a, so far, unexplained heterogeneity across the schemes. Conclusions for policymaking remain abstract.

The relevance of the review findings for policymaking would significantly increase, if they would be complemented with further insights on other policies and outcomes and with information on why some policy schemes are more effective than others. If a policymaker needs to decide what climate policy to choose, comprehensive information on its effects on distributional outcomes, employment, economic competitiveness, innovation effects, leakage, etc. would be required in addition to the mere emissions reduction effect. While traditional literature reviews often cover more than one outcome within one review [140, 305, 306, 308, 325], the more rigorous collection and synthesis of systematic reviews puts limits to what can be achieved within a single review. Findings should also be comparable with similar information on alternative policy options. I do not provide any meaningful comparisons of the effects found for carbon pricing, as I am not aware of any other systematic review on the effectiveness of other climate policies. A comprehensive assessment of climate policy options therefore requires large numbers of systematic reviews across the full range of policies and outcomes.

5.1.3 Realist synthesis leads the way towards an evidence-based theory of how and under what conditions carbon pricing works

The realist synthesis, conducted in Chapter 4, is meant to overcome some of the shortcomings of the aggregative meta-analysis, making it more informative for policymaking. The configurative approach to synthesise the evidence, searching for explanations how and why carbon pricing is or is not effective, aims to answer some of the questions, policymakers would certainly come across when considering what policy to implement in their jurisdiction. The assessment of the mechanisms triggered by the policy in different contextual settings allows to assume, how the same mechanism would unfold, if the policy is implemented in a different context. It also allows to gain an idea, how different actors and societal groups may be affected by the policy measure.

The realist synthesis conducted here aims to derive an evidence-based theory of how and under what conditions carbon pricing works, but the evidence for some of the context-mechanism-outcome configurations is yet rather thin. The review of 80 ex-post policy evaluations provides a considerable amount of relevant evidence, but this is spread across nine hypothesised mechanisms and various country and sector contexts. Some of the hypotheses of how carbon pricing works, can be tested only against very few primary studies. None of the mechanisms can be evaluated in each possible context setting, leaving a number of relevant research gaps, and only few context-mechanism-outcome configurations can be validated across a number of complementary research findings.

Some of the mechanisms reviewed in the realist synthesis were reviewed before, but the synthesis presented here is the first to attempt to organise these into a coherent set of mechanisms leading to emissions reductions. Previous reviews have reviewed evidence on fuel switching [305], low-carbon technology adoption [138–140], investments [139, 305, 306], research and development [139, 140, 305, 306], reduced production [138, 307], and leakage [308]. The evidence is, however, not systematically used to generate theory on the causal chain from the policy implementation to emissions reductions. The realist synthesis approach used here helps to structure all these pieces of evidence into a coherent whole. It also systematically studies each mechanism in varying contexts.

Moving towards evaluations of what it is about a policy that does or does not work, appears to be particularly useful in the case of carbon pricing policies. Views among scientists towards carbon pricing vary considerably [326], to the extent that they are often perceived to be divided into camps of supporters [164, 165] and opponents [169, 171, 172]. I would argue that the opposing views are not all too far apart, but that the one side focuses largely on the upsides and the other side on the downsides of the policy. Using the available empirical evidence to construct an evidence-based theory of the causal mechanisms, potential barriers, and enabling factors, could form a basis for more nuanced scientific discussions on the specific mechanisms that may or may not work in a specific setting. The findings on a given mechanism may also be used to assess to what extent a different climate policy may be able to achieve comparable or better outcomes. Such mechanism-by-mechanism evaluations may also aid the discourse of how policies should best be combined in policy mixes to achieve optimal outcomes.

The presented findings from the realist synthesis are more tangible for policymaking processes, but should be treated with caution. A drawback of the meta-analysis, as stated above, is its narrow answer to a narrow research question. The realist synthesis instead presents multifaceted insights into how carbon pricing works. It is very appealing to use these findings to inform policymaking, but due to the lower amount of considered evidence for each hypothesis, the findings should be considered more preliminary than those derived by the meta-analysis. To increase the robustness of the findings, they need to be further supported with more evidence. For some of the hypotheses, this might be done using an aggregative research method such as a meta-analysis, while for others a configurative method is more suitable, depending on the available evidence. This would also make the findings more reliable to inform policymakers and be used in IPCC assessments.

Realist syntheses vary widely in their comprehensiveness and rigour. The potential of the method to iteratively derive and test hypotheses, reviewing many lines of evidence, leads some research projects to grow very large [59, 60]. Other attempts are more comparable in their scope to the synthesis provided in Chapter 4, limiting the review to a single set of studies identified through one search [129, 327, 328]. We are also not the first to conduct a realist synthesis as a complement to a more aggregative systematic review, sharing the same set of primary studies [329, 330]. Others have, instead of conducting a full realist synthesis, provided sketches how a realist synthesis could be applied in their research area [244, 331, 332]. Compared to these, we go a step further, by providing a full case study of the method applied to carbon pricing policy evaluations. Still, the rigour of our realist synthesis could be improved by critically appraising the reviewed studies, as commonly performed in other applications of the method [59, 127, 128, 329].

The large bandwidth of how realist synthesis is applied in other research fields and the absence such studies in climate policy evaluation, leaves me without a clear guidance to follow. Instead, I directly refer to methodological guidelines [6, 28, 248, 333]. The examples, provided in the methodological literature, are only to a limited extent transferable to climate policy evaluations. The large flexibility in the realist synthesis approach, to be adaptable to heterogenous evidence sources, requires particular care to ensure a high quality of the review. The realist synthesis conducted here, should therefore be seen as an early attempt of this method in the field. More methods work is required to develop tangible guidance for relist synthesis for climate policy evaluations. The same is probably true for configurative reviews in general, which are even more rare in this field than meta-analyses. If no standards are developed, there is a risk that the line between systematic and non-systematic reviews will be too blurry and the established tradition of reviews without clear methods will persist.

5.2 Challenges in the application of evidence synthesis methods

Evidence synthesis is not yet widely applied in the evaluation of climate policies. I want to exploit the experience gathered in this thesis, from the application of three evidence synthesis methods to the field of climate policy evaluation, to identify challenges on the way to mainstreaming evidence synthesis practices in this research field. These challenges range from structural challenges in adjusting research practices and applying new approaches to methodological challenges arising from the adoption of research methods from other disciplines to fit the needs of climate policy evaluation.

5.2.1 Large and growing evidence base

This thesis together with previous assessments of the climate change literature reveal a fundamental evidence synthesis gap within a rapidly growing research field, pointing towards an overwhelming amount of research work required to get on top of all the urgently needed policy insights. Previous studies have shown that climate change research is characterised by an exponential growth in published study findings [14, 15, 320]. Chapter 2 finds that climate policy evaluations are rarely synthesised in systematic reviews or meta-analyses. Within more than 4,000 ex-post carbon pricing studies only a single systematic review is found (with a handful of additional systematic reviews being published after the cut-off date for the map). A similar evidence synthesis gap appears to be present in the broader climate policy literature, when exploring the data by Callaghan *et al.* [15] (with the search terms "systematic review" and "meta-analysis" returning 150 and 99 articles).

The large and growing evidence base not only poses the challenge where to start with synthesising this valuable evidence, but also how to make the evidence synthesis methods fit for purpose to synthesise large amounts of evidence with limited resources. Evidence syntheses are labour intensive research projects, requiring a lot of work for searching and screening the literature, for extracting information and appraising the study quality, before the synthesis can be conducted. This lengthy process often results in systematic reviews being outdated already on the day of their publication [99] – a challenge certainly relevant for the rapidly growing area of climate policy research.

To tackle this challenge, evidence syntheses need to make best use of artificial intelligence [90, 334, 335]. In particular, natural language processing has made great progress over the past few years [145, 163, 336, 337]. As evidence synthesis in large parts is a task of identifying, classifying, and extracting information from texts, natural language processing tools hold a great potential to assist human reviewers with these tasks [88, 90, 92, 335]. In this thesis, I made use of some novel methods for screening and labelling data. During the screening for all three syntheses conducted here, I used machine-learning to predict the relevance of unseen documents, based on the inclusion decisions of the manually screened documents [91, 92, 143]. This allowed to screen documents in the order of their predicted relevance and to stop screening before all documents were seen. In Chapter 2, we also tested how well the evidence and gap map can be updated using machine-learning tools, to facilitate regular updates. Other steps of the review or mapping process could not yet be automated at reasonable levels of reliability [88, 338].

Systematic reviews, to date, have not been automated to the same level as evidence maps [88]. The in-depth assessment of the primary research requires a rigorous extraction of detailed information on the studied population, intervention, outcome, study design, and measured effect size, together with a critical appraisal [23, 24, 339, 340]. Data extraction tools for these tasks are starting to be developed [335, 341, 342], but their application is yet mainly focussed on the eligibility screening and extraction of broader types of information from the abstracts of studies [88]. Future developments of automated methods to reliably extract more types of information from the primary studies, could significantly lower the manual workloads and speed up the process of systematic reviews [335, 338].

Even where machine-learning tools are available, this does not replace all the manual work. As one of the first applications in evidence synthesis, I not only utilised machine-learning to sort literature based on their predicted relevance, but also applied a statistical stopping criterion to evaluate how many of the predicted most likely relevant documents I need to screen before it is safe to stop [144, 343–345]. With currently available methods for prioritisation and stopping, applied on large sets of literature, this still leaves a lot of screening work to human coders. In the screening for Chapter 3 and 4, 4,302 titles and abstracts had to be screened, to find 80 actually relevant studies, before reaching 90% certainty that at least 90% of relevant studies have been identified (based on the stopping criterion by Callaghan and Müller-Hansen [144]). For the mapping in Chapter 2, we stopped screening after 4,054 relevant and 6,061 irrelevant documents had been seen, being 95% confident that we have found at least 90% of the relevant documents (based on the stopping criterion by Callaghan and Müller-Hansen [147]).

The research community will be required to evaluate a number of trade-offs posed by the large amounts of evidence requiring synthesis. While we should aim for rigour and speed at the same time, there are trade-offs between accuracy and time [91]. Stopping criteria in screening processes make this trade-off very transparent. Can we accept recalls of 80%, 90%, 95%, if this allows for substantial time savings? Similar trade-offs have always been part of evidence synthesis, mainly in the literature search. The breadth of the search query, the number of bibliomentric databases, the efforts to trace grey literature, etc. are equally decisive factors to determine how much of the relevant literature we will identify [86, 91]. Similar trade-offs between accuracy and time characterise the mapping approach applied in Chapter 2. We chose to collect information only from abstracts, allowing to capture large amounts of literature in less time, at the risk of missing some information that would only be provided in the full text [90, 321, 322]. Scaling evidence synthesis methods in climate research in a reasonable time, will require a sound balance between accuracy and speed that keeps up with the necessary rigour without loosing too much time.

The rapid growth in the literature can further be tackled by the application of living evidence syntheses, to avoid the syntheses being outdated shortly after their completion. In Chapter 2, we tested how literature maps can be regularly updated using machine-learning tools to create a living repository of the relevant literature. The systematic reviews conducted for this thesis, on the other hand, are static assessments of the evidence at the time of the synthesis. While regular updates would fundamentally increase the value of any evidence synthesis in a time of rapid growth in evidence, research practices, to date, are not favourable for living evidence. PhD theses are a very good example of prevalent research conditions not being aligned with long-term commitments of regularly updating a synthesis for many years. Frequent updates of the syntheses are easier to achieve, the less manual inputs are required. But as a full automation of evidence syntheses is not in sight, challenges with living evidence will remain.

Besides the technological advancements needed to close the evidence synthesis gap of climate policy evaluations, there is also an essential need to grow the community of researchers conducting evidence synthesis on climate policies. With the commencing seventh assessment cycle of the IPCC, it is just the right time for more wide spread synthesis activities in climate and there is some indication that the research community is moving in that direction [108, 346]. In section 5.3.3 below, I sketch how an uptake of evidence syntheses practices could be organised to deliver the best benefits.

5.2.2 Lack of established practice for evidence synthesis in climate research

The novelty of evidence synthesis in climate research necessarily comes with a lack of established practices. While sophisticated methods are available form other research fields, the experience from this thesis shows that the concepts need to be adapted for their application to climate research questions. Such methods refinement is an exciting task to work on, but also comes with additional workload and higher responsibility for the involved researchers. Established standards in the form of review guidelines or, at least, common practices, existing frameworks, and reusable resources, would allow to focus more dedicatedly on the research topic, while the methods would be guided by established concepts. Without such common practices researchers are required to develop and test their methods to ensure an objective assessment of the synthesised evidence.

Across the three syntheses conducted here, I notice an absence of standards from the development of protocols and search queries, all the way to the reporting and publishing of the findings. The established standard of pre-registering evidence syntheses and publishing research protocols, present in other research areas [23, 24], is non-existent in climate policy evaluation. For this thesis, I developed two protocols [135, 188], uploaded to an open data repository, which will likely never be read. There are no standards established yet where and how to search for evidence. A lack of standardised vocabulary and thesaurus impede efficient and effective literature searches [347, 348]. Frameworks and typologies for methods, policies, outcomes, contexts, mechanisms, etc. are not yet regularly available. In Chapter 2, I develop and synthesise typologies of methods and policy outcomes that may form a starting point for more established classification systems. These would help to define and justify eligibility criteria and categorisations in systematic assessments of evidence.

More specific examples of missing guidelines become apparent from the meta-analysis conducted in Chapter 3. Differing disciplinary views complicated the choice of the best method for the averaging of effect sizes [40, 41] and a lack of benchmark studies left us without guidance on how to best harmonise the reported effects or which risk of bias tool to apply in the critical appraisal. The choice in favour of a random effects instead of a fixed effects model for averaging effect sizes (which alleviates the assumption of a convergence to a common effect size mean) [23, 41] resulted in difficulties of applying state of the art publication bias assessments, as advocated by econometricians [40, 349]. For the harmonisation of effect sizes, we decided for an interpretable representation as percentage emission change, instead of the more commonly used correlation coefficients [41], assuming that easily interpretable findings increase the usefulness of the review for policymakers and the general public. In the selection of a risk of bias tool, we decided for a simplified version of existing tools [132, 350], due to limited prior information what types of studies we would expect to identify in our review. In hindsight, we might as well have used an existing tool for quasi-experimental studies [133, 351], as these make up the largest part of the reviewed evidence.

These and other methodological decisions taken for the evidence syntheses, presented in this thesis, may be challenged. The provided syntheses may still contribute to paving the way towards establishing common practices for synthesising evidence from climate policy assessments. In the meantime the reference to guidelines from other research fields [24, 26, 86, 132] and other early syntheses of climate change research (summarised in section 1.2.3) are helpful resources to build on. The exchange with other researchers working on related projects, has revealed to me the great value of exchanging resources in this early stage of methods development, to avoid that each synthesis project starts from scratch.

5.2.3 Comparability of policies across countries

Evidence synthesis may bring a great benefit in learning from climate policy experience across countries, but should not ignore that policy designs and country contexts differ significantly. Any synthesis needs to thoroughly reflect upon which policies to include and how these can be compared.

There is no universal categorisation of which policies are comparable. In this thesis, I synthesise evidence on carbon taxes and cap-and-trade policies, but exclude other taxation instruments, offset-trading, and many other instruments sharing some similarity with the reviewed policies. Similar ambiguity will be present for regulatory instruments, subsidies, bans, or whichever policy instrument one may want to study. This may be further complicated as two climate policies implemented in two jurisdictions may differ in many aspects of their policy design, what emissions and sectors they are applied to, what supplementary policies may already exist, etc. Reviewers are tasked to take a conscious decision which policies should be compared within a review and which ones should be excluded. In the case of carbon pricing, we took advantage of an established list of policies compiled by the World Bank [29], which we used to define which policies we include in our syntheses.

Even after defining a consistent line between inclusion and exclusion, heterogeneity in policy instruments persists. The synthesis method needs to carefully balance between aggregating the evidence to condense knowledge while appreciating the heterogeneity of the studied instruments and contexts. Synthesised evidence needs to reduce the complexity contained in the reviewed primary studies to provide a meaningful overview of their study findings. At the same time, researchers should avoid the risk to overlook important nuance by equalising the evidence across heterogeneous environments.

In this thesis, I, therefore, present all evidence aggregated to the scheme level. The main finding of the meta-analysis, conducted in Chapter 3, could alternatively have been reduced to presenting the average emissions reduction effect across all reviewed schemes. Instead, I chose to emphasise the range of the emissions reductions by scheme (see Figure 3.2), which highlights a significant – and so far insufficiently understood – heterogeneity in the effectiveness of the schemes. Similarly, the realist synthesis in Chapter 4 evaluates all evidence with respect to the scheme it was derived for. The quantitative heterogeneity assessment provided in section 3.3.4 underscores that the scheme is so far the best explanatory factor of the heterogeneity in effects. Still, both systematic reviews assume enough similarity between the studied policies to merit a synthesis of the findings.

Future evidence syntheses of climate policies are advised to make use of the wide range of methodological options to balance between the aggregation of knowledge and an appreciation of heterogeneities between policies and contexts. First, it is important to clearly define which policies to include in a systematic review and which to leave out. Second, the range of synthesis approaches allows to capture both, similarities and heterogeneity in the reviewed evidence. Synthesis methods such as meta-regression and realist synthesis prove particularly useful to study the heterogeneity. These and other configurative research methods should be further exploited to provide a nuanced understanding of climate policies. And even aggregative research approaches, like meta-analysis, provide the flexibility to present synthesised evidence at lower levels of aggregation instead of aggregating all evidence to a single average effect.

5.2.4 Identification challenges in primary studies transmit to synthesis

Evidence synthesis needs to reflect upon the reliability of the primary evidence [23, 24, 130]. While this is the case for any synthesis project, synthesising the evidence from policy evaluations, needs to take into account the identification problems specific to the non-experimental data assessed in primary studies. Policy evaluations are not conducted in laboratories and therefore need to overcome the challenges in the identification of causal effects in real-world data [352, 353].

Policy evaluations need to work with the data available to them (provided in data repositories or collected by the authors). Among the policy evaluations, collected for Chapter 3 and 4, 67% use macro level data on the emissions of a country, region, or sector. Issues with the causal inference from macro level data are well documented. A prominent example from economics is the causal inference to explain the drivers of the gross domestic product [354, 355]. The large number of potential drivers of economic growth challenges the identification of single causal factors [354, 356, 357]. Climate researchers identify an equally long list of factors driving greenhouse gas emissions [358–362], which need to be considered in any attempt to identify the causal effect of a specific factor or policy.

The causal inference of the effect of a given policy on another variable, like emissions, is greatly supported by quasi-experimental study designs, studying variables across time and across study objects, affected and unaffected of the policy, in addition to controlling for other confounding factors [353, 363–366]. Still, these studies need to carefully address the challenges resulting from the non-random policy application and from unobserved differences between the treatment and control group [367, 368]. In addition, there are often multiple policies introduced at around the same time, complicating the separation of the effects of each of the policies [16, 243].

Systematic reviews were introduced to reduce the reliance on single research findings and to systematically compare findings across studies [30, 369, 370]. Synthesising the evidence across many primary studies, tests the consistency of the included research findings and provides a more robust assessment of the studied effects [41, 371].

Critical appraisal of the study designs of each primary study further allows to identify potential biases resulting from the identification challenges listed above [132, 350, 351]. Critical appraisal systematically assesses to what extent a given study is subject to the common sources of bias. For a quasi-experimental study this includes questions on the measurement of the treatment and the outcome, potential endogeneity, the selection of the control group, uncontrolled confounding factors, and the statistical method used [133].

The critical appraisal conducted in Chapter 3 reveals that 46% of the reviewed studies have a medium or high risk of bias, underlining the identification challenges in climate policy evaluations. In particular, the biased studies use non-representative data, struggle to identify control groups with sufficient similarities to the treated study objects, or are confounded by other policy changes happening around the same time.

Systematic reviews can help to gain a robust understanding of climate policies, despite the present identification challenges. Appraising and synthesising the evidence across studies, shows what we can confidently say about a policy and where uncertainties remain. Using multiple lines of evidence further helps to overcome challenges present in specific study designs. We should exploit the potentials of the entire array of aggregative and configurative systematic review methods. Meta-analyses can combine, appraise, and systematically compare the evidence on the causal effect of a policy on its outcomes. Configurative approaches, like realist synthesis, can bring in further lines of evidence, where quantitative policy evaluations fail to establish the causality between policy and outcome. This thesis provides an early attempt to combine these methods for climate policy learning that should be further refined to exploit the full potential of all the valuable evidence by overcoming some of the challenges present in single studies.

Identification challenges in the evaluation of real-world climate policies need to be considered, both, in primary research and in evidence syntheses. Caveats in primary and secondary research need to be discussed and made transparent to achieve continuos and collaborative learning. I openly share the relevant data of this thesis to allow other researchers to challenge and refine how methods are applied in the evidence syntheses conducted here.

5.3 How evidence synthesis can enhance climate policy evaluation for IPCC assessments

Based on the experience gathered throughout this thesis, I am convinced that evidence synthesis has a great potential for accelerating our policy learning in global environmental assessments and beyond. I here want to sketch how evidence syntheses could be effectively used to facilitate the work of the IPCC and how research collaboration could structure the evidence synthesis progress.

5.3.1 Systematic reviews facilitate assessments of large primary evidence in global environmental assessments

A widespread availability of systematic reviews on manifold climate change research questions would allow the IPCC to keep up with the rapid growth of relevant literature, without exponentiating the workloads for the involved authors. While the number of citations in IPCC reports has rapidly increased over the three decades of its existence, its coverage of the climate change literature kept declining due to the even faster growth of the relevant literature [5, 14]. Reliable systematic reviews would allow IPCC authors to assess the available evidence more comprehensively without the need to review ever more primary studies.

Rigorous synthesis conducted in systematic reviews, provides new insights that go beyond a mere summary of the evidence. Their methods are designed to make sense of the reviewed research findings. The meta-analysis, conducted in Chapter 3, for instance, harmonises the effect sizes reported across primary studies, allowing to average and compare them across the policy schemes. It critically appraises each study and assesses the heterogeneity in research findings. Similarly, the realist synthesis, conducted in Chapter 4, tests hypotheses on the emissions reduction pathways across studies, assessing their validity across country and sector contexts. All these analyses provide IPCC authors with additional insights on what conclusions can be drawn from the available evidence, instead of having hundreds or thousands of primary research findings side by side without any systematic comparison of how these support or contradict each other.

Unlike traditional literature reviews, systematic reviews provide a traceable and transparent assessment of the reviewed evidence, allowing IPCC authors to reassess and make their own judgement of the primary evidence. Systematic reviews are transparent on the eligibility criteria for the search and selection of reviewed documents, the appraisal of their research design, and the methods used for synthesis and they make the extracted data publicly available. An IPCC assessment could, based on these information, evaluate to what extent the review provides a comprehensive, reliable, and unbiased answer to the question addressed by the author team. Transparent data reporting even allows the IPCC authors to reassess the provided data with a changed research angle, in case this is relevant to their assessment.

If research questions addressed in the IPCC assessment deviate from those assessed in the systematic review, reassessments can be facilitated by the data collected for the systematic review. Taking the example of the two reviews conducted for this thesis, there are numerous ways in which the assessment may want to adjust the research questions at hand. It might want to change the system boundaries by assessing carbon taxes and cap-and-trade schemes separately or adding additional policy schemes, update with newly published studies, conduct alternative analyses, or alter the appraisal decisions. For the meta-analysis the data and code are publicly available and may feed into alternative assessments to better fit the research questions by the reviewing body. The same holds for the realist synthesis, where the origin of each hypothesis and linked evidence is made transparent in the research report, allowing for a reassessment for subsets or amended with newly added primary studies or a variation of the research question. The collection and preprocessing of the primary evidence facilitates supplementary assessments of the data with drastically reduced workloads for the IPCC author team.

5.3.2 Literature maps provide a common ground for global environmental assessments

Throughout all stages of the IPCC assessment cycle, from the scoping to the approval, an accessible overview of what literature is out there can facilitate the discussions between all parties on what evidence the IPCC can, should, and does assess. From the definition of the assessment scope, through several rounds of assessment and review, up until the final approval of the report, assessments are guided by both the interest in policy relevant research questions as well as the availability of research addressing these questions. A structured and interactive tool to navigate within the available literature, provides a low-barrier device to understand which questions are studied in the literature and which ones might not be assessable due to gaps in the evidence. This may be of particular value to facilitate the transdisciplinary discussions involving researchers, governments, external experts and observer organisations.

Evidence and gap maps with the detail provided in the map compiled in Chapter 2, extended to all relevant strands of the literature, would facilitate such discussions. This would require detailed maps of the relevant research, including climate policy evaluations, drivers of climate change, impacts, technological potentials, political economy factors, etc., and across all types of evidence – ex-post and ex-ante, qualitative and quantitative.

At the assessment stage, a well structured map eases the access for IPCC authors and reduces the efforts needed to trace the available literature. With the large amount of relevant evidence, to date, even domain experts struggle to keep an overview of all the available evidence. Climate policy questions are addressed across multiple scientific disciplines and published in hundreds of different journals. A systematic map of the available evidence lowers the risk of overlooking relevant branches of the literature. An evidence and gap map, as presented in this thesis, may point the author team directly towards systematic reviews of parts of the evidence, which lighten their assessment task.

While the map may serve as a valuable common resource, it should not be prescriptive of what research questions can and should be addressed in an assessment or not. The IPCC process involves many experts with knowledge of additional relevant literature or an informed view, why other literature may be less relevant. The evidence map may miss studies for a number of reasons. The evidence and gap map, compiled in Chapter 2, for instance, is based on three bibliometric databases with a focus on academic literature and is limited to English language literature. Evidence that is published, for instance, in government reports or in other languages will be missing in this map. The map would also not resolve the increasingly raised demand to more thoroughly integrate indigenous knowledge into the assessment reports [372–374]. Additional literature searches should therefore always be conducted to broaden the scope of the considered evidence.

5.3.3 Research community should work towards ecosystems of evidence syntheses on all relevant climate policy questions

The two sections above suggest how evidence syntheses could be used for IPCC and other global environmental assessments. An uptake of evidence synthesis practices, however, needs to arise mainly outside of these assessments, driven by the research community. Based on the experience from this thesis, I want to reflect on some ideas how such work could be fostered to advance the science-policy interface.

A vision for evidence syntheses on the full range of policy relevant questions would consist of large numbers of interlinked syntheses guided by policy needs and the availability of high quality evidence. Key elements of such an *ecosystem of evidence syntheses* would be evidence maps for scoping the evidence and defining research questions, expert elicitation to construct typologies of policies, mechanisms, contexts, outcomes, and research methods, leading towards coordinated and interlinked systematic reviews across policies and outcomes using supplementary systematic review methods.

Starting from large scale evidence maps, as provided by Callaghan *et al.* [15] and Sietsma *et al.* [109], the climate policy research landscape could be partitioned and ecosystems of systematic reviews and more fine-grained maps be developed for each branch of the map. Following the typology by Callaghan *et al.* [15] partitioning of the evidence base could follow the lines of policies or sectors. As conducted in

Chapter 2, more fine grained evidence and gap maps for each policy could reveal the state of evidence on the specific policy and show where systematic reviews are yet conducted or missing, informing the further research agenda.

Systematic reviews studying each outcome for each policy should be coordinated to the extent that comparable frameworks for the policies, outcomes, methods, and context factors are applied and unnecessary duplication of work is avoided. Common understanding of these frameworks may be facilitated by consulting the evidence maps compiled in preparation of the review ecosystem as well as consultations with users of the evidence. The reviews conducted for this thesis show that the application of different synthesis methods also merit overlaps between the evidence assessed in multiple reviews, such that coordination and data sharing among review teams can avoid duplicated work for searching and identifying relevant literature. Similarly, the findings of one systematic review could point towards research questions that should be assessed by supplementary systematic reviews.

Illustrating the idea using the evidence syntheses compiled for this thesis, an ecosystem of carbon pricing reviews could be kicked off. The evidence and gap map identifies evidence synthesis gaps for many of the policy outcomes, such as environmental effectiveness, leakage, innovation, distributional effects, public perception, or health co-benefits. While Chapter 3 provides a systematic review of environmental effectiveness to fill the respective gap, each of the other outcomes would merit a dedicated systematic review. The map could provide a starting point for the identification of relevant studies for each of the reviews, reducing their workloads and facilitating an agreement of system boundaries between reviews. As mentioned above, a group of researchers has started to fill some of the evidence synthesis gaps identified by the map.

While coordinating between research teams may be beneficial, it should also be explored how systematic reviews with different system boundaries can be communicated to policymakers. The systematic reviews conducted here, for instance, focus on ex-post assessments of both carbon taxes and emissions trading schemes, while other reviews may cover both ex-post and ex-ante studies or focus exclusively on carbon taxes. All these reviews may complement each other, but their distinction also need to be well understood.

In an ideal world all the interlinked evidence syntheses should be living [99], such that any policy advice could always build upon the most up-to-date evidence. While limited resources may not allow for each systematic review to be constantly updated, living evidence and gap maps would be an essential requirement to become the central element of the proposed ecosystem of evidence syntheses. Reviewers could then, at any point in time, use the map as the starting point to identify an evidence synthesis gap and extract the relevant primary evidence for their review.

To conclude, ecosystems of evidence syntheses could be comprehensive, interlinked assessments of the entire climate policy evidence. These could provide a structured overview of the state of knowledge on what climate policies work, under what conditions, why, causing which side effects or co-benefits. They could inform policy decisions as well as environmental assessments by synthesising all the valuable evidence we already have to address the climate crisis.

References

[1] UNFCCC, *Adoption of the Paris Agreement (FCCC/CP/L.9/Rev.1)*. Geneva (Switzerland): United Nations Framework Convention on Climate Change, United Nations Office, 2015.

[2] IPCC, *Climate Change 2022: Mitigation of Climate Change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, P. Shukla, J. Skea, R. Slade, A. A. Khourdajie, R. van Diemen, D. McCollum, M. Pathak, S. Some, P. Vyas, R. Fradera, M. Belkacemi, A. Hasija, G. Lisboa, S. Luz, and J. Malley, Eds. Cambridge, UK and New York, NY, USA: Cambridge University Press, 2022. DOI: 10.1017/9781009157926.

[3] United Nations Environment Programme, A. Olhoff, C. Bataille, J. Christensen, M. Den Elzen, T. Fransen, N. Grant, K. Blok, J. Kejun, E. Soubeyran, W. Lamb, K. Levin, J. Portugal-Pereira, M. Pathak, T. Kuramochi, C. Strinati, S. Roe, and J. Rogelj, *Emissions Gap Report 2024: No more hot air ... please! With a massive gap between rhetoric and reality, countries draft new climate commitments*. United Nations Environment Programme, 2024. DOI: 10.59117/20.500.11822/46404.

[4] L. Berrang–Ford, F. Döbbe, R. Garside, N. Haddaway, W. F. Lamb, J. C. Minx, W. Viechtbauer, V. Welch, and H. White, "Editorial: Evidence synthesis for accelerated learning on climate solutions," *Campbell Systematic Reviews*, vol. 16, no. 4, p. 1, 2020. DOI: 10.1002/cl2.1128.

[5] J. C. Minx, M. Callaghan, W. F. Lamb, J. Garard, and O. Edenhofer, "Learning about climate change solutions in the IPCC and beyond," *Environmental Science & Policy*, vol. 77, pp. 252–259, 2017. DOI: 10.1016/j.envsci.2017.05.014.

[6] R. Pawson, T. Greenhalgh, G. Harvey, and K. Walshe, "Realist synthesis: An introduction," *Manchester: ESRC Research Methods Programme Working Paper Series, University of Manchester*, 2004.

[7] G. Iacobuta, N. K. Dubash, P. Upadhyaya, M. Deribe, and N. Höhne, "National climate change mitigation legislation, strategy and targets: A global update," *Climate Policy*, vol. 18, no. 9, pp. 1114–1132, 2018. DOI: 10.1080/14693062.2018.1489772.

[8] S. Schaub, J. Tosun, A. Jordan, and J. Enguer, "Climate policy ambition: Exploring a policy density perspective," *Politics and Governance*, vol. 10, no. 3, 2022. DOI: 10.17645/pag.v10i3.5347.

[9] N. K. Dubash, C. Mitchell, E. L. Boasson, M. J. Borbor-Córdova, S. Fifita, E. Haites, M. Jaccard, F. Jotzo, S. Naidoo, P. Romero-Lankao, W. Shen, M. Shlapak, and L. Wu, "National and sub-national policies and institutions," in *Climate Change 2022 - Mitigation of Climate Change*, IPCC, Ed., 1st ed., Cambridge University Press, 2023, pp. 1355–1450. DOI: 10.1017/9781009157926.015.

[10] D. Nachtigall, L. Lutz, M. Cárdenas Rodríguez, I. Haščič, and R. Pizarro, "The climate actions and policies measurement framework: A structured and harmonised climate policy database to monitor countries' mitigation action," OECD Environment Working Papers 203, 2022, Series: OECD Environment Working Papers Volume: 203. DOI: 10.1787/2caa60ce-en.

[11] F. M. D'Arcangelo, T. Kruse, and M. Pisu, "Identifying and tracking climate change mitigation strategies with a cluster-based assessment," *npj Climate Action*, vol. 3, no. 1, p. 86, 2024. DOI: 10.1038/s44168-024-00158-6.

[12] NewClimate Institute, Wageningen University and Research, and PBL Netherlands Environmental Assessment Agency, *Climate policy database*, version v.2023, 2024. DOI: 10.5281/ZENODO.10869734.

[13] Grantham Research Institute at the London School of Economics and Climate Policy Radar. "Climate change laws of the world." (2024), [Online]. Available: https://climate-laws.org (visited on 11/13/2024).

[14] M. Callaghan, J. C. Minx, and P. M. Forster, "A topography of climate change research," *Nature Climate Change*, vol. 10, no. 2, pp. 118–123, 2020. DOI: 10.1038/s41558-019-0684-5.

[15] M. Callaghan, L. Banisch, N. Doebbeling-Hildebrandt, D. Edmondson, C. Flachsland, W. Lamb, S. Levi, F. Müller-Hansen, E. Posada, S. Vasudevan, and J. Minx, *Mapping climate mitigation policy literature using machine learning: Disparities between scientific attention, policy density, and emissions*, 2024. DOI: 10.21203/rs.3.rs-3817176/v1.

[16] A. Stechemesser, N. Koch, E. Mark, E. Dilger, P. Klösel, L. Menicacci, D. Nachtigall, F. Pretis, N. Ritter, M. Schwarz, H. Vossen, and A. Wenzel, "Climate policies that achieved major emission reductions: Global evidence from two decades," *Science*, vol. 385, no. 6711, pp. 884–892, 2024. DOI: 10.1126/science.adl6547.

[17] S. M. S. U. Eskander and S. Fankhauser, "Reduction in greenhouse gas emissions from national climate legislation," *Nature Climate Change*, vol. 10, no. 8, pp. 750–756, 2020. DOI: 10.1038/s41558-020-0831-z.

[18] R. Rafaty, G. Dolphin, and F. Pretis, "Carbon pricing and the elasticity of CO2 emissions," *Institute for New Economic Thinking Working Paper Series*, pp. 1–84, 2020. DOI: 10.36687/inetwp140.

[19] N. R. Haddaway, A. Bethel, L. V. Dicks, J. Koricheva, B. Macura, G. Petrokofsky, A. S. Pullin, S. Savilaakso, and G. B. Stewart, "Eight problems with

literature reviews and how to fix them," *Nature Ecology & Evolution*, vol. 4, no. 12, pp. 1582–1589, 2020. DOI: 10.1038/s41559-020-01295-x.

[20] N. Haddaway, P. Woodcock, B. Macura, and A. Collins, "Making literature reviews more reliable through application of lessons from systematic reviews," *Conservation Biology*, vol. 29, no. 6, pp. 1596–1605, 2015. DOI: 10.1111/cobi.12541.

[21] M. Egger and G. D. Smith, "Meta-analysis: Potentials and promise," *BMJ*, vol. 315, no. 7119, pp. 1371–1374, 1997. DOI: 10.1136/bmj.315.7119.1371.

[22] Campbell Collaboration, "Campbell systematic reviews: Policies and guidelines," vol. 1.8, 2021.

[23] J. Higgins, J. Thomas, and J. Chandler, *Cochrane handbook for systematic reviews of interventions* (Cochrane book series), Second edition. 2019.

[24] Collaboration for Environmental Evidence, *Guidelines and Standards for Evidence Synthesis in Environmental Management*, A. S. Pullin, G. K. Frampton, B. Livoreil, and G. Petrokofsky, Eds. 2022, vol. Version 5.1.

[25] K. L. James, N. P. Randall, and N. R. Haddaway, "A methodology for systematic mapping in environmental sciences," *Environmental Evidence*, vol. 5, no. 1, p. 245, 2016. DOI: 10.1186/s13750-016-0059-6.

[26] A. Saran and H. White, "Evidence and gap maps: A comparison of different approaches," *Campbell Systematic Reviews*, vol. 14, no. 1, pp. 1–38, 2018. DOI: 10.4073/cmdp.2018.2.

[27] M. Kastner, J. Antony, C. Soobiah, S. E. Straus, and A. C. Tricco, "Conceptual recommendations for selecting the most appropriate knowledge synthesis method to answer research questions related to complex evidence," *Journal of clinical epidemiology*, vol. 73, pp. 43–49, 2016. DOI: 10.1016/j.jclinepi.2015.11.022.

[28] R. Pawson, T. Greenhalgh, G. Harvey, and K. Walshe, "Realist review-a new method of systematic review designed for complex policy interventions," *Journal of health services research & policy*, vol. 10, pp. 21–34, Suppl. 1 2005.

[29] World Bank. "Carbon pricing dashboard: Key statistics for 2023 on initiatives implemented." (2023), [Online]. Available: https://carbonpricingdashboard.worldbank.org/map_data (visited on 08/14/2023).

[30] G. V. Glass, "Primary, secondary, and meta-analysis of research," *Educational Researcher*, vol. 5, no. 10, pp. 3–8, 1976. DOI: 10.3102/0013189X005010003.

[31] G. V. Glass, "Integrating findings: The meta-analysis of research," *Review of Research in Education*, vol. 5, p. 351, 1977. DOI: 10.2307/1167179.

[32] M. L. Smith and G. V. Glass, "Meta-analysis of psychotherapy outcome studies.," *American Psychologist*, vol. 32, no. 9, pp. 752–760, 1977. DOI: 10.1037/0003-066X.32.9.752.

[33] M. W. Lipsey and D. B. Wilson, "The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis.," *American Psychologist*, vol. 48, no. 12, pp. 1181–1209, 1993. DOI: 10.1037/0003-066X.48.12.1181.

[34] R. E. Slavin, "Meta-analysis in education: How has it been used?" *Educational Researcher*, vol. 13, no. 8, pp. 6–15, 1984. DOI: 10.3102/0013189X013008006.

[35] T. D. Spector and S. G. Thompson, "The potential and limitations of meta-analysis.," *Journal of Epidemiology & Community Health*, vol. 45, no. 2, pp. 89–92, 1991. DOI: 10.1136/jech.45.2.89.

[36] L. Bero and R. Drummond, "The cochrane collaboration: Preparing, maintaining, and disseminating systematic reviews of the effects of health care," *JAMA*, vol. 274, no. 24, p. 1935, 1995. DOI: 10.1001/jama.1995.03530240045039.

[37] J. H. Littell and H. White, “The campbell collaboration: Providing better evidence for a better world,” *Research on Social Work Practice*, vol. 28, no. 1, pp. 6–12, 2018. DOI: 10.1177/1049731517703748.

[38] A. Petrosino, R. F. Boruch, C. Rounding, S. McDonald, and I. Chalmers, “The campbell collaboration social, psychological, educational and criminological trials register (c2-SPECTR) to facilitate the preparation and maintenance of systematic reviews of social and educational interventions,” *Evaluation & Research in Education*, vol. 14, no. 3, pp. 206–219, 2000. DOI: 10.1080/09500790008666973.

[39] T. Stanley, H. Doucouliagos, M. Giles, J. H. Heckemeyer, R. J. Johnston, P. Laroche, J. P. Nelson, M. Paldam, J. Poot, G. Pugh, R. S. Rosenberger, and K. Rost, “Meta-analysis of economics research reporting guidelines,” *Journal of Economic Surveys*, vol. 27, no. 2, pp. 390–394, 2013. DOI: 10.1111/joes.12008.

[40] Z. Irsova, H. Doucouliagos, T. Havranek, and T. D. Stanley, “Meta-analysis of social science research: A practitioner’s guide,” *Journal of Economic Surveys*, vol. 38, no. 5, pp. 1547–1566, 2024. DOI: 10.1111/joes.12595.

[41] E. Ringquist, *Meta-analysis for public management and policy*. John Wiley & Sons, 2013.

[42] S. J. Cooke, C. N. Cook, V. M. Nguyen, J. C. Walsh, N. Young, C. Cvitanovic, M. J. Grainger, N. P. Randall, M. Muir, A. N. Kadykalo, K. A. Monk, and A. S. Pullin, “Environmental evidence in action: On the science and practice of evidence synthesis and evidence-based decision-making,” *Environmental Evidence*, vol. 12, no. 1, p. 10, 2023. DOI: 10.1186/s13750-023-00302-5.

[43] W. J. Sutherland, A. S. Pullin, P. M. Dolman, and T. M. Knight, “The need for evidence-based conservation,” *Trends in Ecology & Evolution*, vol. 19, no. 6, pp. 305–308, 2004. DOI: 10.1016/j.tree.2004.03.018.

[44] J. Noyes, J. Popay, A. Pearson, K. Hannes, and A. Booth, "Qualitative research and cochrane reviews," in *Cochrane Handbook for Systematic Reviews of Interventions*, J. P. Higgins and S. Green, Eds., 1st ed., Wiley, 2008, pp. 571–591. DOI: 10.1002/9780470712184.ch20.

[45] J. Noyes, A. Booth, M. Cargo, K. Flemming, R. Garside, K. Hannes, A. Harden, J. Harris, S. Lewin, T. Pantoja, and J. Thomas, "Cochrane qualitative and implementation methods group guidance series—paper 1: Introduction," *Journal of Clinical Epidemiology*, vol. 97, pp. 35–38, 2018. DOI: 10.1016/j.jclinepi.2017.09.025.

[46] J. Noyes, A. Booth, M. Cargo, K. Flemming, A. Harden, J. Harris, R. Garside, K. Hannes, T. Pantoja, and J. Thomas, "Chapter 4: Searching for and selecting studies [last updated september 2024]," in *Cochrane Handbook for Systematic Reviews of Interventions*, J. P. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, and V. A. Welch, Eds., 6.5, 2024.

[47] J. L. Harris, A. Booth, M. Cargo, K. Hannes, A. Harden, K. Flemming, R. Garside, T. Pantoja, J. Thomas, and J. Noyes, "Cochrane qualitative and implementation methods group guidance series—paper 2: Methods for question formulation, searching, and protocol development for qualitative evidence synthesis," *Journal of Clinical Epidemiology*, vol. 97, pp. 39–48, 2018. DOI: 10.1016/j.jclinepi.2017.10.023.

[48] J. Noyes, A. Booth, K. Flemming, R. Garside, A. Harden, S. Lewin, T. Pantoja, K. Hannes, M. Cargo, and J. Thomas, "Cochrane qualitative and implementation methods group guidance series—paper 3: Methods for assessing methodological limitations, data extraction and synthesis, and confidence in synthesized qualitative findings," *Journal of Clinical Epidemiology*, vol. 97, pp. 49–58, 2018. DOI: 10.1016/j.jclinepi.2017.06.020.

[49] A. Harden, J. Thomas, M. Cargo, J. Harris, T. Pantoja, K. Flemming, A. Booth, R. Garside, K. Hannes, and J. Noyes, "Cochrane qualitative and implementation methods group guidance series—paper 5: Methods for integrat-

ing qualitative and implementation evidence within intervention effectiveness reviews," *Journal of Clinical Epidemiology*, vol. 97, pp. 70–78, 2018. DOI: 10.1016/j.jclinepi.2017.11.029.

[50] D. Gough, J. Thomas, and S. Oliver, "Clarifying differences between review designs and methods," *Systematic Reviews*, vol. 1, no. 1, p. 28, 2012. DOI: 10.1186/2046-4053-1-28.

[51] V. Entwistle, D. Firnigl, M. Ryan, J. Francis, and P. Kinghorn, "Which experiences of health care delivery matter to service users and why? a critical interpretive synthesis and conceptual map," *Journal of Health Services Research & Policy*, vol. 17, no. 2, pp. 70–78, 2012. DOI: 10.1258/jhsrp.2011.011029.

[52] S. Farrelly and H. Lester, "Therapeutic relationships between mental health service users with psychotic disorders and their clinicians: A critical interpretive synthesis," *Health & Social Care in the Community*, vol. 22, no. 5, pp. 449–460, 2014. DOI: 10.1111/hsc.12090.

[53] M. J. Hammer, E. A. Ercolano, F. Wright, V. V. Dickson, D. Chyun, and G. D. Melkus, "Self-management for adult patients with cancer: An integrative review," *Cancer Nursing*, vol. 38, no. 2, E10–E26, 2015. DOI: 10.1097/NCC.0000000000000122.

[54] D. McKechnie, J. Pryor, and M. J. Fisher, "Falls and fallers in traumatic brain injury (TBI) rehabilitation settings: An integrative review," *Disability and Rehabilitation*, vol. 37, no. 24, pp. 2291–2299, 2015. DOI: 10.3109/09638288.2014.1002578.

[55] T. P. Jones, M. C. Katapodi, and J. S. Lockhart, "Factors influencing breast cancer screening and risk assessment among young african american women: An integrative review of the literature," *Journal of the American Association of Nurse Practitioners*, vol. 27, no. 9, pp. 521–529, 2015. DOI: 10.1002/2327-6924.12223.

[56] E. Taylor, C. McKevitt, and F. Jones, "Factors shaping the delivery of acute inpatient stroke therapy: A narrative synthesis," *Journal of Rehabilitation Medicine*, vol. 47, no. 2, pp. 107–119, 2015. DOI: 10.2340/16501977-1918.

[57] V. Guise, J. Anderson, and S. Wiig, "Patient safety risks associated with telecare: A systematic review and narrative synthesis of the literature," *BMC Health Services Research*, vol. 14, no. 1, p. 588, 2014. DOI: 10.1186/s12913-014-0588-z.

[58] K. Loudon, N. Santesso, M. Callaghan, J. Thornton, J. Harbour, K. Graham, R. Harbour, I. Kunnamo, H. Liira, E. McFarlane, K. Ritchie, and S. Treweek, "Patient and public attitudes to and awareness of clinical practice guidelines: A systematic review with thematic and narrative syntheses," *BMC Health Services Research*, vol. 14, no. 1, p. 321, 2014. DOI: 10.1186/1472-6963-14-321.

[59] J. Greenhalgh, S. Dalkin, K. Gooding, E. Gibbons, J. Wright, D. Meads, N. Black, J. M. Valderas, and R. Pawson, "Functionality and feedback: A realist synthesis of the collation, interpretation and utilisation of patient-reported outcome measures data to improve patient care," *Health Services and Delivery Research*, vol. 5, no. 2, pp. 1–280, 2017. DOI: 10.3310/hsdr05020.

[60] M. S. Barnish, S. Y. Tan, S. Robinson, A. Taeihagh, and G. Melendez-Torres, "A realist synthesis to develop an explanatory model of how policy instruments impact child and maternal health outcomes," *Social Science & Medicine*, vol. 339, p. 116 402, 2023. DOI: 10.1016/j.socscimed.2023.116402.

[61] G. Wong, T. Greenhalgh, and R. Pawson, "Internet-based medical education: A realist review of what works, for whom and in what circumstances," *BMC Medical Education*, vol. 10, no. 1, p. 12, 2010. DOI: 10.1186/1472-6920-10-12.

[62] M. Röing and M. Sanner, "A meta-ethnographic synthesis on phenomenographic studies of patients' experiences of chronic illness," *International Jour-*

nal of Qualitative Studies on Health and Well-being, vol. 10, no. 1, p. 26 279, 2015. DOI: 10.3402/qhw.v10.26279.

[63] I. Tombor, L. Shahab, A. Herbec, J. Neale, S. Michie, and R. West, “Smoker identity and its potential role in young adults’ smoking behavior: A meta-ethnography.,” *Health Psychology*, vol. 34, no. 10, pp. 992–1003, 2015. DOI: 10.1037/hea0000191.

[64] A. Papadopoulos, J. M. Sargeant, S. E. Majowicz, B. Sheldrick, C. McKeen, J. Wilson, and C. E. Dewey, “Enhancing public trust in the food safety regulatory system,” *Health Policy*, vol. 107, no. 1, pp. 98–103, 2012. DOI: 10.1016/j.healthpol.2012.05.010.

[65] R. Arnold and D. Fletcher, “A research synthesis and taxonomic classification of the organizational stressors encountered by sport performers,” *Journal of Sport and Exercise Psychology*, vol. 34, no. 3, pp. 397–429, 2012. DOI: 10.1123/jsep.34.3.397.

[66] A. Conway, V. Schadewaldt, R. Clark, C. Ski, D. R. Thompson, and L. Doering, “The psychological experiences of adult heart transplant recipients: A systematic review and meta-summary of qualitative findings,” *Heart & Lung*, vol. 42, no. 6, pp. 449–455, 2013. DOI: 10.1016/j.hrtlng.2013.08.003.

[67] E. Asonganyi, M. Vaghasia, C. Rodrigues, A. Phadtare, A. Ford, R. Pietrobon, J. Atashili, and C. Lynch, “Factors affecting compliance with clinical practice guidelines for pap smear screening among healthcare providers in africa: Systematic review and meta-summary of 2045 individuals,” *PLoS ONE*, vol. 8, no. 9, A. N. Cheung, Ed., e72712, 2013. DOI: 10.1371/journal.pone.0072712.

[68] A. Edwards, N. Pang, V. Shiu, and C. Chan, “Review: The understanding of spirituality and the potential role of spiritual care in end-of-life and palliative care: A meta-study of qualitative research,” *Palliative Medicine*, vol. 24, no. 8, pp. 753–770, 2010. DOI: 10.1177/0269216310375860.

[69] J. Smith, V. Swallow, and I. Coyne, "Involving parents in managing their child's long-term condition—a concept synthesis of family-centered care and partnership-in-care," *Journal of Pediatric Nursing*, vol. 30, no. 1, pp. 143–159, 2015. DOI: 10.1016/j.pedn.2014.10.014.

[70] J. Midtgaard, N. M. Hammer, C. Andersen, A. Larsen, D.-M. Bruun, and M. Jarden, "Cancer survivors' experience of exercise-based cancer rehabilitation – a meta-synthesis of qualitative research," *Acta Oncologica*, vol. 54, no. 5, pp. 609–617, 2015. DOI: 10.3109/0284186X.2014.995777.

[71] T. Sandsdalen, R. Hov, S. Høye, I. Rystedt, and B. Wilde-Larsson, "Patients' preferences in palliative care: A systematic mixed studies review," *Palliative Medicine*, vol. 29, no. 5, pp. 399–419, 2015. DOI: 10.1177/0269216314557882.

[72] J. P. Domecq, G. Prutsky, T. Elraiyah, Z. Wang, M. Nabhan, N. Shippee, J. P. Brito, K. Boehmer, R. Hasan, B. Firwana, P. Erwin, D. Eton, J. Sloan, V. Montori, N. Asi, A. M. Abu Dabrh, and M. H. Murad, "Patient engagement in research: A systematic review," *BMC Health Services Research*, vol. 14, no. 1, p. 89, 2014. DOI: 10.1186/1472-6963-14-89.

[73] P. A. Collins and M. V. Hayes, "The role of urban municipal governments in reducing health inequities: A meta-narrative mapping analysis," *International Journal for Equity in Health*, vol. 9, no. 1, p. 13, 2010. DOI: 10.1186/1475-9276-9-13.

[74] K. Fierz, D. Nicca, and R. Spirig, "Perceived HIV symptom manageability: Synthesis of a new use for a known concept," *Journal of Advanced Nursing*, vol. 69, no. 1, pp. 229–241, 2013. DOI: 10.1111/j.1365-2648.2012.06068.x.

[75] N. R. Haddaway, K. Hedlund, L. E. Jackson, T. Kätterer, E. Lugato, I. K. Thomsen, H. B. Jørgensen, and P.-E. Isberg, "How does tillage intensity affect soil organic carbon? a systematic review," *Environmental Evidence*, vol. 6, no. 1, p. 30, 2017. DOI: 10.1186/s13750-017-0108-9.

[76] R. Maharaj, I. Raffaele, and J. Wendon, "Rapid response systems: A systematic review and meta-analysis," *Critical Care*, vol. 19, no. 1, p. 254, 2015. DOI: 10.1186/s13054-015-0973-y.

[77] T. M. Khanna, G. Baiocchi, M. Callaghan, F. Creutzig, H. Guias, N. R. Haddaway, L. Hirth, A. Javaid, N. Koch, S. Laukemper, A. Löschel, M. d. M. Zamora Dominguez, and J. C. Minx, "A multi-country meta-analysis on the role of behavioural change in reducing energy consumption and CO2 emissions in residential buildings," *Nature Energy*, vol. 6, no. 9, pp. 925–932, 2021. DOI: 10.1038/s41560-021-00866-x.

[78] G. Filippini, C. Del Giovane, L. Vacchi, R. D'Amico, C. Di Pietrantonj, D. Beecher, and G. Salanti, "Immunomodulators and immunosuppressants for multiple sclerosis: A network meta-analysis," *Cochrane Database of Systematic Reviews*, vol. 2013, no. 6, Cochrane Multiple Sclerosis and Rare Diseases of the CNS Group, Ed., 2013. DOI: 10.1002/14651858.CD008933.pub2.

[79] T. A. Furukawa, K. Shinohara, E. Sahker, E. Karyotaki, C. Miguel, M. Ciharova, C. L. Bockting, J. J. Breedvelt, A. Tajika, H. Imai, E. G. Ostinelli, M. Sakata, R. Toyomoto, S. Kishimoto, M. Ito, Y. Furukawa, A. Cipriani, S. D. Hollon, and P. Cuijpers, "Initial treatment choices to achieve sustained response in major depression: A systematic review and network meta-analysis," *World Psychiatry*, vol. 20, no. 3, pp. 387–396, 2021. DOI: 10.1002/wps.20906.

[80] K. Flemming, A. Booth, R. Garside, Ö. Tunçalp, and J. Noyes, "Qualitative evidence synthesis for complex interventions and guideline development: Clarification of the purpose, designs and relevant methods," *BMJ Global Health*, vol. 4, e000882, Suppl 1 2019. DOI: 10.1136/bmjgh-2018-000882.

[81] D. L. Katz, A.-L. Williams, C. Girard, J. Goodman, B. Comerford, A. Behrman, and M. B. Bracken, "The evidence base for complementary and alternative medicine: Methods of evidence mapping with application to CAM," *Alternative therapies in health and medicine*, no. 9, pp. 22–37, 2003.

[82] B. Snilstveit, M. Vojtkova, A. Bhavsar, and M. Gaarder, *Evidence gap maps—a tool for promoting evidence-informed policy and prioritizing future research*. Policy research working paper, the World Bank, 2013.

[83] J. Mingers and L. Leydesdorff, "A review of theory and practice in scientometrics," *European Journal of Operational Research*, vol. 246, no. 1, pp. 1–19, 2015. DOI: 10.1016/j.ejor.2015.04.002.

[84] I. D. Cooper, "Bibliometrics basics," *Journal of the Medical Library Association : JMLA*, vol. 103, no. 4, pp. 217–218, 2015. DOI: 10.3163/1536-5050.103.4.013.

[85] N. R. Haddaway, C. Bernes, B.-G. Jonsson, and K. Hedlund, "The benefits of systematic mapping to evidence-based environmental management," *Ambio*, vol. 45, no. 5, pp. 613–620, 2016. DOI: 10.1007/s13280-016-0773-x.

[86] B. Snilstveit, M. Vojtkova, A. Bhavsar, J. Stevenson, and M. Gaarder, "Evidence & gap maps: A tool for promoting evidence informed policy and strategic research agendas," *Journal of Clinical Epidemiology*, vol. 79, pp. 120–129, 2016. DOI: 10.1016/j.jclinepi.2016.05.015.

[87] M. C. McKinnon, S. H. Cheng, R. Garside, Y. J. Masuda, and D. C. Miller, "Sustainability: Map the evidence," *Nature*, vol. 528, no. 7581, pp. 185–187, 2015. DOI: 10.1038/528185a.

[88] L. Schmidt, A. N. Finnerty Mutlu, R. Elmore, B. K. Olorisade, J. Thomas, and J. P. T. Higgins, "Data extraction methods for systematic review (semi)-automation: Update of a living systematic review," *F1000Research*, vol. 10, p. 401, 2023. DOI: 10.12688/f1000research.51117.2.

[89] M. Callaghan, "Machine reading the science of climate change: Computational tools to support evidence-based decision-making in the age of big literature," Ph.D. dissertation, University of Leeds, Leeds, 2021.

[90] N. R. Haddaway, M. Callaghan, A. M. Collins, W. F. Lamb, J. C. Minx, J. Thomas, and D. John, "On the use of computer–assistance to facilitate

systematic mapping," *Campbell Systematic Reviews*, vol. 16, no. 4, p. 9, 2020. DOI: 10.1002/cl2.1129.

[91] A. O'Mara-Eves, J. Thomas, J. McNaught, M. Miwa, and S. Ananiadou, "Using text mining for study identification in systematic reviews: A systematic review of current approaches," *Systematic Reviews*, vol. 4, no. 1, p. 5, 2015. DOI: 10.1186/2046-4053-4-5.

[92] I. Shemilt, A. Simon, G. J. Hollands, T. M. Marteau, D. Ogilvie, A. O'Mara-Eves, M. P. Kelly, and J. Thomas, "Pinpointing needles in giant haystacks: Use of text mining to reduce impractical screening workload in extremely large scoping reviews," *Research Synthesis Methods*, vol. 5, no. 1, pp. 31–49, 2014. DOI: 10.1002/jrsm.1093.

[93] M. Miwa, J. Thomas, A. O'Mara-Eves, and S. Ananiadou, "Reducing systematic review workload through certainty-based screening," *Journal of Biomedical Informatics*, vol. 51, pp. 242–253, 2014. DOI: 10.1016/j.jbi.2014.06.005.

[94] J. H. Elliott *et al.*, "Living systematic review: 1. introduction—the why, what, when, and how," *Journal of Clinical Epidemiology*, vol. 91, pp. 23–30, 2017. DOI: 10.1016/j.jclinepi.2017.08.010.

[95] J. H. Elliott, T. Turner, O. Clavisi, J. Thomas, J. P. T. Higgins, C. Mavergames, and R. L. Gruen, "Living systematic reviews: An emerging opportunity to narrow the evidence-practice gap," *PLoS Medicine*, vol. 11, no. 2, e1001603, 2014. DOI: 10.1371/journal.pmed.1001603.

[96] J. Thomas *et al.*, "Living systematic reviews: 2. combining human and machine effort," *Journal of Clinical Epidemiology*, vol. 91, pp. 31–37, 2017. DOI: 10.1016/j.jclinepi.2017.08.011.

[97] I. Shemilt, G. Hollands, C. Vigurs, C. Stansfield, and J. Thomas, "Misinformation in COVID-19: A living map of research evidence. [archived final

version, december 2022],” *London: EPPI Centre, UCL Social Research Institute, University College London*, 2022.

[98] EPPI Centre. “Living map of systematic reviews of social sciences research evidence on COVID-19.” (2024), [Online]. Available: https://eppi.ioe.ac.uk/cms/Default.aspx?tabid=3806 (visited on 07/16/2024).

[99] J. Elliott, R. Lawrence, J. C. Minx, O. T. Oladapo, P. Ravaud, B. Tendal Jeppesen, J. Thomas, T. Turner, P. O. Vandvik, and J. M. Grimshaw, “Decision makers need constantly updated evidence synthesis,” *Nature*, vol. 600, no. 7889, pp. 383–385, 2021. DOI: 10.1038/d41586-021-03690-1.

[100] IPCC. “About the IPCC.” (2024), [Online]. Available: https://www.ipcc.ch/about/ (visited on 11/19/2024).

[101] E. Byers, V. Krey, E. Kriegler, K. Riahi, R. Schaeffer, J. Kikstra, R. Lamboll, Z. Nicholls, M. Sandstad, C. Smith, K. van der Wijst, A. Al -Khourdajie, F. Lecocq, J. Portugal-Pereira, Y. Saheb, A. Stromman, H. Winkler, C. Auer, E. Brutschin, M. Gidden, P. Hackstock, M. Harmsen, D. Huppmann, P. Kolp, C. Lepault, J. Lewis, G. Marangoni, E. Müller-Casseres, R. Skeie, M. Werning, K. Calvin, P. Forster, C. Guivarch, T. Hasegawa, M. Meinshausen, G. Peters, J. Rogelj, B. Samset, J. Steinberger, M. Tavoni, and D. van Vuuren, *AR6 scenarios database*, version 1.1, 2022. DOI: 10.5281/ZENODO.5886911.

[102] “Mitigation pathways compatible with long-term goals,” in *Climate Change 2022 - Mitigation of Climate Change*, IPCC, K. Riahi, R. Schaeffer, J. Arango, K. Calvin, C. Guivarch, T. Hasegawa, K. Jiang, E. Kriegler, R. Matthews, G. P. Peters, A. Rao, S. Robertson, A. M. Sebbit, J. Steinberger, M. Tavoni, and D. P. van Vuuren, Eds., 1st ed., Cambridge University Press, 2023, pp. 295–408. DOI: 10.1017/9781009157926.005.

[103] P. N. Edwards, “Global comprehensive models in politics and policymaking,” *Climatic Change*, vol. 32, no. 2, pp. 149–161, 1996.

[104] B. Cointe, C. Cassen, and A. Nadaï, "Organising policy-relevant knowledge for climate action," *Science & Technology Studies*, vol. 32, no. 4, pp. 36–57, 2019. DOI: 10.23987/sts.65031.

[105] IPCC, Ed., *Climate change 2001: mitigation: contribution of Working Group III to the third assessment report of the Intergovernmental Panel on Climate Change*, Cambridge ; New York: Cambridge University Press, 2001, 752 pp.

[106] IPCC, Ed., *Climate change 2007: Mitigation of climate change. Contribution of Working Group III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge: Cambridge university press, 2007.

[107] IPCC, Ed., *Climate Change 2014: Mitigation of climate change Working Group III contribution to the fifth assessment report of the Intergovernmental Panel on Climate Change*, New York: Cambridge university press, 2014.

[108] H. Pearson, "What's the best way to tackle climate change? an 'evidence bank' could help scientists find answers," *Nature*, vol. 630, no. 8017, pp. 540–541, 2024. DOI: 10.1038/d41586-024-01683-4.

[109] A. J. Sietsma, J. D. Ford, M. W. Callaghan, and J. C. Minx, "Progress in climate change adaptation research," *Environmental Research Letters*, vol. 16, no. 5, p. 054 038, 2021. DOI: 10.1088/1748-9326/abf7f3.

[110] W. F. Lamb, M. W. Callaghan, F. Creutzig, R. Khosla, and J. C. Minx, "The literature landscape on 1.5°c climate change and cities," *1.5°C Climate change and urban areas*, vol. 30, pp. 26–34, 2018. DOI: 10.1016/j.cosust.2018.02.008.

[111] M. Sethi, W. Lamb, J. Minx, and F. Creutzig, "Climate change mitigation in cities: A systematic scoping of case studies," *Environmental Research Letters*, vol. 15, no. 9, p. 093 008, 2020. DOI: 10.1088/1748-9326/ab99ff.

[112] W. F. Lamb, F. Creutzig, M. Callaghan, and J. C. Minx, "Learning about urban climate solutions from case studies," *Nature Climate Change*, vol. 9, no. 4, pp. 279–287, 2019. DOI: 10.1038/s41558-019-0440-x.

[113] S. Montfort, M. Callaghan, F. Creutzig, W. Lamb, C. Lu, T. Repke, K. Ge, and J. Minx, *A global systematic map and database of climate change research on cities – insufficient research covers fast-growing cities*, 2024. DOI: 10.21203/rs.3.rs-5092456/v1.

[114] L. Berrang-Ford, A. J. Sietsma, M. Callaghan, J. C. Minx, P. F. D. Scheelbeek, N. R. Haddaway, A. Haines, and A. D. Dangour, "Systematic mapping of global research on climate and health: A machine learning review," *The Lancet Planetary Health*, vol. 5, no. 8, e514–e525, 2021. DOI: 10.1016/S2542-5196(21)00179-0.

[115] M. Berretta, C. Zamawe, P. J. Ferraro, N. Haddaway, J. Minx, B. Snilstveit, and J. Eyers, "Mapping energy efficiency interventions," International Initiative for Impact Evaluation (3ie), 2021, Edition: 2021. DOI: 10.23846/EGM017.

[116] C. G. Parrao, C. Yavuz, Z. Ravat, M. Berretta, D. Rodrigues, M. Bhattacharyya, A. Floridi, B. Sovacool, A. Bethel, Q. Reifmesser, F. Gaved, S. Pilato, S. Asfaw, and B. Snilstveit, "Promoting sustainable energy development through access, renewables and efficient technologies: An evidence gap map," International Initiative for Impact Evaluation (3ie), 2024, Edition: 2024. DOI: 10.23846/EGM032.

[117] N. Ohlendorf, M. Jakob, J. C. Minx, C. Schröder, and J. C. Steckel, "Distributional impacts of carbon pricing: A meta-analysis," *Environmental and Resource Economics*, vol. 78, no. 1, pp. 1–42, 2020. DOI: 10.1007/s10640-020-00521-1.

[118] C. Midões, J. Van Den Bergh, and I. Savin, "A meta-analysis of synergy between carbon pricing and renewable-energy policies," *SSRN Electronic Journal*, 2024. DOI: 10.2139/ssrn.4763473.

[119] F. Mohammadzadeh Valencia, C. Mohren, A. Ramakrishnan, M. Merchert, J. C. Minx, and J. C. Steckel, "Public support for carbon pricing policies and revenue recycling options: A systematic review and meta-analysis of the survey literature," *npj Climate Action*, vol. 3, no. 1, p. 74, 2024. DOI: 10.1038/s44168-024-00153-x.

[120] J. Barrez, "Public acceptability of carbon pricing: Unravelling the impact of revenue recycling," *Climate Policy*, vol. 24, no. 10, pp. 1323–1345, 2024. DOI: 10.1080/14693062.2024.2376747.

[121] S. Cuevas, D. Nachtigall, A. Aguilar Jaber, K. Belesova, J. Falconer, A. Haines, T. Reynolds, T. M. Schuster, S. Whitmee, and R. Green, "Health co-benefits and trade-offs of carbon pricing: A narrative synthesis," *Climate Policy*, pp. 1–19, 2024. DOI: 10.1080/14693062.2024.2356822.

[122] J. Magnetti, G. Dominioni, and B. Gordijn, "Ethics of carbon pricing – a review of the literature," *Climate Policy*, pp. 1–20, 2024. DOI: 10.1080/14693062.2024.2416493.

[123] J. D. Ford, L. Berrang-Ford, and J. Paterson, "A systematic review of observed climate change adaptation in developed nations: A letter," *Climatic Change*, vol. 106, no. 2, pp. 327–336, 2011. DOI: 10.1007/s10584-011-0045-5.

[124] G. Owen, "What makes climate change adaptation effective? a systematic review of the literature," *Global Environmental Change*, vol. 62, p. 102 071, 2020. DOI: 10.1016/j.gloenvcha.2020.102071.

[125] A. Méjean, P. Collins-Sowah, C. Guivarch, F. Piontek, B. Soergel, and N. Taconet, "Climate change impacts increase economic inequality: Evidence from a systematic literature review," *Environmental Research Letters*, vol. 19, no. 4, p. 043 003, 2024. DOI: 10.1088/1748-9326/ad376e.

[126] T. M. Khanna, D. Danilenko, M. Andor, M. Callaghan, J. H. Elliott, T. Repke, L. A. Smith, J. Sanchez, T. V. Bhumika, and J. C. Minx, "Protocol:

Behavioral, information and monetary interventions to reduce energy consumption in households: A "living" systematic review," *Campbell Systematic Reviews*, vol. 20, no. 3, e1424, 2024. DOI: 10.1002/cl2.1424.

[127] N. Willand, I. Ridley, and C. Maller, "Towards explaining the health impacts of residential energy efficiency interventions – a realist review. part 1: Pathways," *Social Science & Medicine*, vol. 133, pp. 191–201, 2015. DOI: 10.1016/j.socscimed.2015.02.005.

[128] N. Willand, C. Maller, and I. Ridley, "Understanding the contextual influences of the health outcomes of residential energy efficiency interventions: Realist review," *Housing Studies*, vol. 35, no. 1, pp. 1–28, 2017. DOI: 10.1080/02673037.2017.1363874.

[129] L. Camprubí, D. Malmusi, R. Mehdipanah, L. Palència, A. Molnar, C. Muntaner, and C. Borrell, "Façade insulation retrofitting policy implementation process and its effects on health equity determinants: A realist review," *Energy Policy*, vol. 91, pp. 304–314, 2016. DOI: 10.1016/j.enpol.2016.01.016.

[130] H. White, B. Albers, M. Gaarder, H. Kornør, J. Littell, Z. Marshall, C. Mathew, T. Pigott, B. Snilstveit, H. Waddington, and V. Welch, "Guidance for producing a campbell evidence and gap map," *Campbell Systematic Reviews*, vol. 16, no. 4, e1125, 2020. DOI: 10.1002/cl2.1125.

[131] 3ie. "Evidence gap maps - evidence hub." (2024), [Online]. Available: https://3ieimpact.org/evidence-hub/evidence-gap-maps (visited on 12/11/2024).

[132] J. A. Sterne, M. A. Hernán, A. McAleenan, B. C. Reeves, and J. P. Higgins, "Assessing risk of bias in a non-randomized study," in *Cochrane Handbook for Systematic Reviews of Interventions*, J. P. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, and V. A. Welch, Eds., 1st ed., Wiley, 2019, pp. 621–641. DOI: 10.1002/9781119536604.ch25.

[133] T. H. Barker, N. Habibi, E. Aromataris, J. C. Stone, J. Leonardi-Bee, K. Sears, S. Hasanoff, M. Klugar, C. Tufanaru, S. Moola, and Z. Munn, "The

revised JBI critical appraisal tool for the assessment of risk of bias for quasi-experimental studies," *JBI Evidence Synthesis*, vol. 22, no. 3, pp. 378–388, 2024. DOI: 10.11124/JBIES-23-00268.

[134] K. Flemming and J. Noyes, "Qualitative evidence synthesis: Where are we at?" *International Journal of Qualitative Methods*, vol. 20, 2021. DOI: 10.1177/1609406921993276.

[135] N. Döbbeling-Hildebrandt, J. Minx, and T. Repke, "Protocol: Carbon pricing outcomes – an evidence and gap map of the ex-post literature," 2024, Publisher: Open Science Framework. DOI: 10.17605/OSF.IO/JWB2R.

[136] S. Kugley, A. Wade, J. Thomas, Q. Mahood, A.-M. K. Jørgensen, K. Hammerstrøm, and N. Sathe, "Searching for studies: A guide to information retrieval for campbell systematic reviews," *Campbell Systematic Reviews*, vol. 13, no. 1, pp. 1–73, 2017. DOI: 10.4073/cmg.2016.1.

[137] J. F. Green, "Does carbon pricing reduce emissions? a review of ex-post analyses," *Environmental Research Letters*, vol. 16, no. 4, p. 043 004, 2021. DOI: 10.1088/1748-9326/abdae9.

[138] A. Köppl and M. Schratzenstaller, "Carbon taxation: A review of the empirical literature," *Journal of Economic Surveys*, joes.12531, 2022. DOI: 10.1111/joes.12531.

[139] J. Lilliestam, A. Patt, and G. Bersalli, "The effect of carbon pricing on technological change for full energy decarbonization: A review of empirical ex-post evidence," *WIREs Climate Change*, vol. 12, no. 1, 2020. DOI: 10.1002/wcc.681.

[140] C. Peñasco, L. D. Anadón, and E. Verdolini, "Systematic review of the outcomes and trade-offs of ten types of decarbonization policy instruments," *Nature Climate Change*, vol. 11, no. 3, pp. 257–265, 2021. DOI: 10.1038/s41558-020-00971-x.

[141] T. Repke and M. Callaghan, "NACSOS-nexus: NLP assisted classification, synthesis and online screening with new and extended usage scenarios," 2024, Publisher: arXiv. DOI: 10.48550/ARXIV.2405.04621.

[142] N. R. Haddaway, B. Macura, P. Whaley, and A. S. Pullin, "ROSES RepOrting standards for systematic evidence syntheses: Pro forma, flow-diagram and descriptive summary of the plan and conduct of environmental systematic reviews and systematic maps," *Environmental Evidence*, vol. 7, no. 1, p. 7, 2018. DOI: 10.1186/s13750-018-0121-7.

[143] R. van de Schoot, J. de Bruin, R. Schram, P. Zahedi, J. de Boer, F. Weijdema, B. Kramer, M. Huijts, M. Hoogerwerf, G. Ferdinands, A. Harkema, J. Willemsen, Y. Ma, Q. Fang, S. Hindriks, L. Tummers, and D. L. Oberski, "An open source machine learning framework for efficient and transparent systematic reviews," *Nature Machine Intelligence*, vol. 3, no. 2, pp. 125–133, 2021. DOI: 10.1038/s42256-020-00287-7.

[144] M. Callaghan and F. Müller-Hansen, "Statistical stopping criteria for automated screening in systematic reviews," *Systematic reviews*, vol. 9, no. 1, p. 273, 2020. DOI: 10.1186/s13643-020-01521-4.

[145] N. Webersinke, M. Kraus, J. A. Bingler, and M. Leippold, *ClimateBert: A pretrained language model for climate-related text*, Version Number: 3, 2021. DOI: 10.48550/ARXIV.2110.12010.

[146] A. Cohan, S. Feldman, I. Beltagy, D. Downey, and D. S. Weld, *SPECTER: Document-level representation learning using citation-informed transformers*, Version Number: 4, 2020. DOI: 10.48550/ARXIV.2004.07180.

[147] M. Callaghan and F. Müller-Hansen. "Calculate reliable stopping criteria for users of machine-learning prioritised screening in systematic reviews." (2024), [Online]. Available: https://mcallaghan.github.io/buscarR/ (visited on 06/03/2024).

[148] A. Fog, "Calculation methods for wallenius' noncentral hypergeometric distribution," *Communications in Statistics - Simulation and Computation*, vol. 37, no. 2, pp. 258–273, 2008. DOI: 10.1080/03610910701790269.

[149] World Bank, *State and Trends of Carbon Pricing 2023*. World Bank, Washington, DC, 2023.

[150] P. Konidari and D. Mavrakis, "A multi-criteria evaluation method for climate change mitigation policy instruments," *Energy Policy*, vol. 35, no. 12, pp. 6235–6257, 2007. DOI: 10.1016/j.enpol.2007.07.007.

[151] W. F. Lamb, M. Antal, K. Bohnenberger, L. I. Brand-Correa, F. Müller-Hansen, M. Jakob, J. C. Minx, K. Raiser, L. Williams, and B. K. Sovacool, "What are the social outcomes of climate policies? a systematic map and review of the ex-post literature," *Environmental Research Letters*, vol. 15, no. 11, p. 113 006, 2020. DOI: 10.1088/1748-9326/abc11f.

[152] L. Neij and K. Åstrand, "Outcome indicators for the evaluation of energy policy instruments and technical change," *Energy Policy*, vol. 34, no. 17, pp. 2662–2676, 2006. DOI: 10.1016/j.enpol.2005.03.012.

[153] F. Schneider and A. F. Wagner, "Tradeable permits and climate change policy," 2002, Publisher: Citeseer.

[154] B. Zhu, J. Tang, P. Wang, and L. Zhang, "Exploring the drivers of carbon market risk: A meta regression analysis," *Journal of Cleaner Production*, vol. 352, p. 131 538, 2022. DOI: 10.1016/j.jclepro.2022.131538.

[155] E. A. Akl *et al.*, "Living systematic reviews: 4. living guideline recommendations," *Journal of Clinical Epidemiology*, vol. 91, pp. 47–53, 2017. DOI: 10.1016/j.jclinepi.2017.08.009.

[156] M. Simmonds *et al.*, "Living systematic reviews: 3. statistical methods for updating meta-analyses," *Journal of Clinical Epidemiology*, vol. 91, pp. 38–46, 2017. DOI: 10.1016/j.jclinepi.2017.08.008.

[157] M. Kowarsch, J. Jabbour, C. Flachsland, M. T. J. Kok, R. Watson, P. M. Haas, J. C. Minx, J. Alcamo, J. Garard, P. Riousset, L. Pintér, C. Langford, Y. Yamineva, C. von Stechow, J. O'Reilly, and O. Edenhofer, "A road map for global environmental assessments," *Nature Climate Change*, vol. 7, no. 6, pp. 379–382, 2017. DOI: 10.1038/nclimate3307.

[158] B. C. O'Leary, K. Kvist, H. R. Bayliss, G. Derroire, J. R. Healey, K. Hughes, F. Kleinschroth, M. Sciberras, P. Woodcock, and A. S. Pullin, "The reliability of evidence review methodology in environmental science and conservation," *Environmental Science & Policy*, vol. 64, pp. 75–82, 2016. DOI: 10.1016/j.envsci.2016.06.012.

[159] T. Stanley, "Wheat from chaff: Meta-analysis as quantitative literature review," *Journal of Economic Perspectives*, vol. 15, no. 3, pp. 131–150, 2001. DOI: 10.1257/jep.15.3.131.

[160] I. Chalmers, M. B. Bracken, B. Djulbegovic, S. Garattini, J. Grant, A. M. Gülmezoglu, D. W. Howells, J. P. A. Ioannidis, and S. Oliver, "How to increase value and reduce waste when research priorities are set," *The Lancet*, vol. 383, no. 9912, pp. 156–165, 2014. DOI: 10.1016/S0140-6736(13)62229-1.

[161] A. Alshami, M. Elsayed, E. Ali, A. E. E. Eltoukhy, and T. Zayed, "Harnessing the power of ChatGPT for automating systematic review process: Methodology, case study, limitations, and future directions," *Systems*, vol. 11, no. 7, p. 351, 2023. DOI: 10.3390/systems11070351.

[162] S. Wang, H. Scells, S. Zhuang, M. Potthast, B. Koopman, and G. Zuccon, *Zero-shot generative large language models for systematic review screening automation*, Version Number: 2, 2024. DOI: 10.48550/ARXIV.2401.06320.

[163] I. Beltagy, K. Lo, and A. Cohan, "SciBERT: A pretrained language model for scientific text," 2019, Publisher: arXiv Version Number: 3. DOI: 10.48550/ARXIV.1903.10676.

[164] Climate Leadership Council. "Economists' statement on carbon dividends organized by the climate leadership council." (2019), [Online]. Available: https://www.econstatement.org (visited on 03/30/2022).

[165] High-Level Commission on Carbon Prices, *Report of the High-Level Commission on Carbon Prices*. Washington, DC: World Bank, 2017.

[166] W. J. Baumol and W. E. Oates, "The use of standards and prices for protection of the environment," in *The Economics of Environment*, P. Bohm and A. V. Kneese, Eds., London: Palgrave Macmillan UK, 1971, pp. 53–65. DOI: 10.1007/978-1-349-01379-1_4.

[167] W. Montgomery, "Markets in licenses and efficient pollution control programs," *Journal of Economic Theory*, vol. 5, no. 3, pp. 395–418, 1972. DOI: 10.1016/0022-0531(72)90049-X.

[168] T. Sterner, E. B. Barbier, I. Bateman, I. van den Bijgaart, A.-S. Crépin, O. Edenhofer, C. Fischer, W. Habla, J. Hassler, O. Johansson-Stenman, A. Lange, S. Polasky, J. Rockström, H. G. Smith, W. Steffen, G. Wagner, J. E. Wilen, F. Alpízar, C. Azar, D. Carless, C. Chávez, J. Coria, G. Engström, S. C. Jagers, G. Köhlin, Å. Löfgren, H. Pleijel, and A. Robinson, "Policy design for the anthropocene," *Nature Sustainability*, vol. 2, no. 1, pp. 14–21, 2019. DOI: 10.1038/s41893-018-0194-x.

[169] D. Rosenbloom, J. Markard, F. W. Geels, and L. Fuenfschilling, "Why carbon pricing is not sufficient to mitigate climate change—and how "sustainability transition policy" can help," *Proceedings of the National Academy of Sciences*, vol. 117, no. 16, pp. 8664–8668, 2020. DOI: 10.1073/pnas.2004093117.

[170] I. Savin, S. Drews, S. Maestre-Andrés, and J. Van Den Bergh, "Public views on carbon taxation and its fairness: A computational-linguistics analysis," *Climatic Change*, vol. 162, no. 4, pp. 2107–2138, 2020. DOI: 10.1007/s10584-020-02842-y.

[171] A. Patt and J. Lilliestam, "The case against carbon prices," *Joule*, vol. 2, no. 12, pp. 2494–2498, 2018. DOI: 10.1016/j.joule.2018.11.018.

[172] J. F. Green, "Beyond carbon pricing: Tax reform is climate policy," *Global Policy*, vol. 12, no. 3, pp. 372–379, 2021. DOI: 10.1111/1758-5899.12920.

[173] L. Rotaris and R. Danielis, "The willingness to pay for a carbon tax in italy," *Transportation Research Part D: Transport and Environment*, vol. 67, pp. 659–673, 2019. DOI: 10.1016/j.trd.2019.01.001.

[174] Y. Hu, S. Ren, Y. Wang, and X. Chen, "Can carbon emission trading scheme achieve energy conservation and emission reduction? evidence from the industrial sector in china," *Energy Economics*, vol. 85, p. 104 590, 2020. DOI: 10.1016/j.eneco.2019.104590.

[175] M. Leroutier, "Carbon pricing and power sector decarbonization: Evidence from the UK," *Journal of Environmental Economics and Management*, vol. 111, p. 102 580, 2022. DOI: 10.1016/j.jeem.2021.102580.

[176] N. Rivers and B. Schaufele, "Salience of carbon taxes in the gasoline market," *Journal of Environmental Economics and Management*, vol. 74, pp. 23–36, 2015. DOI: 10.1016/j.jeem.2015.07.002.

[177] D. Xiang and C. Lawley, "The impact of british columbia's carbon tax on residential natural gas consumption," *Energy Economics*, vol. 80, pp. 206–218, 2019. DOI: 10.1016/j.eneco.2018.12.004.

[178] K. Gugler, A. Haxhimusa, and M. Liebensteiner, "Effectiveness of climate policies: Carbon pricing vs. subsidizing renewables," *Journal of Environmental Economics and Management*, vol. 106, p. 102 405, 2021. DOI: 10.1016/j.jeem.2020.102405.

[179] M. Espey, "Gasoline demand revisited: An international meta-analysis of elasticities," *Energy Economics*, vol. 20, no. 3, pp. 273–295, 1998. DOI: 10.1016/S0140-9883(97)00013-3.

[180] M. Brons, P. Nijkamp, E. Pels, and P. Rietveld, "A meta-analysis of the price elasticity of gasoline demand. a SUR approach," *Energy Economics*, vol. 30, no. 5, pp. 2105–2122, 2008. DOI: 10.1016/j.eneco.2007.08.004.

[181] T. Havranek, Z. Irsova, and K. Janda, "Demand for gasoline is more price-inelastic than commonly thought," *Energy Economics*, vol. 34, no. 1, pp. 201–207, 2012. DOI: 10.1016/j.eneco.2011.09.003.

[182] J. A. Espey and M. Espey, "Turning on the lights: A meta-analysis of residential electricity demand elasticities," *Journal of Agricultural and Applied Economics*, vol. 36, no. 1, pp. 65–81, 2004. DOI: 10.1017/S1074070800021866.

[183] X. Labandeira, J. M. Labeaga, and X. López-Otero, "A meta-analysis on the price elasticity of energy demand," *Energy Policy*, vol. 102, pp. 549–568, 2017. DOI: 10.1016/j.enpol.2017.01.002.

[184] E. Haites, "Carbon taxes and greenhouse gas emissions trading systems: What have we learned?" *Climate Policy*, vol. 18, no. 8, pp. 955–966, 2018. DOI: 10.1080/14693062.2018.1492897.

[185] J. Van Den Bergh and I. Savin, "Impact of carbon pricing on low-carbon innovation and deep decarbonisation: Controversies and path forward," *Environmental and Resource Economics*, vol. 80, no. 4, pp. 705–715, 2021. DOI: 10.1007/s10640-021-00594-6.

[186] J. C. Minx, N. R. Haddaway, and K. L. Ebi, "Planetary health as a laboratory for enhanced evidence synthesis," *The Lancet Planetary Health*, vol. 3, no. 11, e443–e445, 2019.

[187] Collaboration for Environmental Evidence, *Guidelines and Standards for Evidence Synthesis in Environmental Management*, A. S. Pullin, G. K. Frampton, B. Livoreil, and G. Petrokofsky, Eds. 2018, vol. Version 5.0.

[188] N. Döbbeling, K. Miersch, J. Minx, T. Khanna, M. Callaghan, N. Koch, J. C. Steckel, N. Ohlendorf, M. Bachelet, and O. Edenhofer, "Protocol: Effectiveness of carbon pricing – a systematic review and meta-analysis of the ex-post

literature," 2022, Publisher: Open Science Framework. DOI: 10.17605/OSF.IO/854VP.

[189] M. Callaghan, F. Müller-Hansen, J. Hilaire, and Y. T. Lee, *NACSOS: NLP assisted classification, synthesis and online screening*, version v0.1.0, 2020. DOI: 10.5281/ZENODO.4121526.

[190] N. Haddaway, B. Macura, P. Whaley, and A. Pullin, "ROSES flow diagram for systematic reviews. version 1.0," 2018. DOI: 10.6084/M9.FIGSHARE.5897389.

[191] W. Viechtbauer, "Bias and efficiency of meta-analytic variance estimators in the random-effects model," *Journal of Educational and Behavioral Statistics*, vol. 30, no. 3, pp. 261–293, 2005. DOI: 10.3102/10769986030003261.

[192] W. Viechtbauer, "Conducting meta-analyses in *R* with the metafor package," *Journal of Statistical Software*, vol. 36, no. 3, 2010. DOI: 10.18637/jss.v036.i03.

[193] E. Tipton, "Small sample adjustments for robust variance estimation with meta-regression.," *Psychological Methods*, vol. 20, no. 3, pp. 375–393, 2015. DOI: 10.1037/met0000011.

[194] J. E. Pustejovsky and E. Tipton, "Small-sample methods for cluster-robust variance estimation and hypothesis testing in fixed effects models," *Journal of Business & Economic Statistics*, vol. 36, no. 4, pp. 672–683, 2018. DOI: 10.1080/07350015.2016.1247004.

[195] M. Egger, G. D. Smith, M. Schneider, and C. Minder, "Bias in meta-analysis detected by a simple, graphical test," *BMJ*, vol. 315, no. 7109, pp. 629–634, 1997. DOI: 10.1136/bmj.315.7109.629.

[196] T. D. Stanley, "Meta-regression methods for detecting and estimating empirical effects in the presence of publication selection," *Oxford Bulletin of Economics and Statistics*, 2007. DOI: 10.1111/j.1468-0084.2007.00487.x.

[197] T. D. Stanley, H. Doucouliagos, and J. P. A. Ioannidis, "Finding the power to reduce publication bias: Finding the power to reduce publication bias," *Statistics in Medicine*, 2017. DOI: 10.1002/sim.7228.

[198] J. P. A. Ioannidis, T. D. Stanley, and H. Doucouliagos, "The power of bias in economics research," *The Economic Journal*, vol. 127, no. 605, F236–F265, 2017. DOI: 10.1111/ecoj.12461.

[199] J. Cohen, "Some statistical issues in psychological research," *Handbook of clinical psychology*, pp. 95–121, 1965.

[200] World Bank. "Carbon pricing dashboard: Key statistics for 2022 on initiatives implemented." (2022), [Online]. Available: https://carbonpricingdashboard.worldbank.org/map_data (visited on 07/11/2022).

[201] International Carbon Action Partnership. "Allowance price explorer." (2022), [Online]. Available: https://icapcarbonaction.com/en/ets-prices (visited on 02/12/2022).

[202] A. E. Raftery, D. Madigan, and J. A. Hoeting, "Bayesian model averaging for linear regression models," *Journal of the American Statistical Association*, vol. 92, no. 437, pp. 179–191, 1997. DOI: 10.1080/01621459.1997.10473615.

[203] J. Bajzik, T. Havranek, Z. Irsova, and J. Schwarz, "Estimating the armington elasticity: The importance of study design and publication bias," *Journal of International Economics*, vol. 127, p. 103 383, 2020. DOI: 10.1016/j.jinteco.2020.103383.

[204] T. Havranek, M. Rusnak, and A. Sokolova, "Habit formation in consumption: A meta-analysis," *European Economic Review*, vol. 95, pp. 142–167, 2017. DOI: 10.1016/j.euroecorev.2017.03.009.

[205] S. Zeugner and M. Feldkircher, "Bayesian model averaging employing fixed and flexible priors: The BMS package for *R*," *Journal of Statistical Software*, vol. 68, no. 4, 2015. DOI: 10.18637/jss.v068.i04.

[206] T. S. Eicher, C. Papageorgiou, and A. E. Raftery, "Default priors and predictive performance in bayesian model averaging, with application to growth determinants," *Journal of Applied Econometrics*, vol. 26, no. 1, pp. 30–55, 2011. DOI: 10.1002/jae.1112.

[207] E. Ley and M. F. Steel, "On the effect of prior assumptions in bayesian model averaging with applications to growth regression," *Journal of Applied Econometrics*, vol. 24, no. 4, pp. 651–674, 2009. DOI: 10.1002/jae.1057.

[208] C. Fernández, E. Ley, and M. F. Steel, "Benchmark priors for bayesian model averaging," *Journal of Econometrics*, vol. 100, no. 2, pp. 381–427, 2001. DOI: 10.1016/S0304-4076(00)00076-2.

[209] J. P. A. Ioannidis, "Why most published research findings are false," *PLoS medicine*, vol. 2, no. 8, e124, 2005. DOI: 10.1371/journal.pmed.0020124.

[210] R. Rosenthal, "The file drawer problem and tolerance for null results," *Psychological Bulletin*, vol. 86, no. 3, pp. 638–641, 1979. DOI: 10.1037/0033-2909.86.3.638.

[211] A. Brodeur, M. Lé, M. Sangnier, and Y. Zylberberg, "Star wars: The empirics strike back," *American Economic Journal: Applied Economics*, vol. 8, no. 1, pp. 1–32, 2016. DOI: 10.1257/app.20150044.

[212] M. L. Weitzman, "Prices vs. quantities," *The Review of Economic Studies*, vol. 41, no. 4, p. 477, 1974. DOI: 10.2307/2296698.

[213] C. Hepburn, "Regulation by prices, quantities, or both: A review of instrument choice," *Oxford Review of Economic Policy*, vol. 22, no. 2, pp. 226–247, 2006. DOI: 10.1093/oxrep/grj014.

[214] M. Kalkuhl and O. Edenhofer, "Prices vs. quantities and the intertemporal dynamics of the climate rent," *SSRN Electronic Journal*, 2010. DOI: 10.2139/ssrn.1605112.

[215] L. H. Goulder and A. R. Schein, "Carbon taxes versus cap and trade: A critical review," *Climate Change Economics*, vol. 04, no. 3, p. 1 350 010, 2013. DOI: 10.1142/S2010007813500103.

[216] J. Foramitti, I. Savin, and J. C. van den Bergh, "Emission tax vs. permit trading under bounded rationality and dynamic markets," *Energy Policy*, vol. 148, p. 112 009, 2021. DOI: 10.1016/j.enpol.2020.112009.

[217] C. Flachsland, S. Brunner, O. Edenhofer, and F. Creutzig, "Climate policies for road transport revisited (II): Closing the policy gap with cap-and-trade," *Energy Policy*, vol. 39, no. 4, pp. 2100–2110, 2011. DOI: 10.1016/j.enpol.2011.01.053.

[218] F. Kesicki and N. Strachan, "Marginal abatement cost (MAC) curves: Confronting theory and practice," *Environmental Science & Policy*, vol. 14, no. 8, pp. 1195–1204, 2011. DOI: 10.1016/j.envsci.2011.08.004.

[219] B.-J. Tang, C.-J. Ji, Y.-J. Hu, J.-X. Tan, and X.-Y. Wang, "Optimal carbon allowance price in china's carbon emission trading system: Perspective from the multi-sectoral marginal abatement cost," *Journal of Cleaner Production*, vol. 253, p. 119 945, 2020. DOI: 10.1016/j.jclepro.2019.119945.

[220] N. Kaufman, A. R. Barron, W. Krawczyk, P. Marsters, and H. McJeon, "A near-term to net zero alternative to the social cost of carbon for setting carbon prices," *Nature Climate Change*, vol. 10, no. 11, pp. 1010–1014, 2020. DOI: 10.1038/s41558-020-0880-3.

[221] J. Strefler, E. Kriegler, N. Bauer, G. Luderer, R. C. Pietzcker, A. Giannousakis, and O. Edenhofer, "Alternative carbon price trajectories can avoid excessive carbon removal," *Nature Communications*, vol. 12, no. 1, p. 2264, 2021. DOI: 10.1038/s41467-021-22211-2.

[222] W. J. McKibbin, R. Shackleton, and P. J. Wilcoxen, "What to expect from an international system of tradable permits for carbon emissions," *Resource and*

Energy Economics, vol. 21, no. 3, pp. 319–346, 1999. DOI: 10.1016/S0928-7655(99)00007-X.

[223] C. Flachsland, R. Marschinski, and O. Edenhofer, "To link or not to link: Benefits and disadvantages of linking cap-and-trade systems," *Climate Policy*, vol. 9, no. 4, pp. 358–372, 2009. DOI: 10.3763/cpol.2009.0626.

[224] T. Sterner, "Fuel taxes: An important instrument for climate policy," *Energy Policy*, vol. 35, no. 6, pp. 3194–3202, 2007. DOI: 10.1016/j.enpol.2006.10.025.

[225] C. A. Dahl, "Measuring global gasoline and diesel price and income elasticities," *Modeling Transport (Energy) Demand and Policies*, vol. 41, pp. 2–13, 2012. DOI: 10.1016/j.enpol.2010.11.055.

[226] J. Morris, S. Paltsev, and J. Reilly, "Marginal abatement costs and marginal welfare costs for greenhouse gas emissions reductions: Results from the EPPA model," *Environmental Modeling & Assessment*, vol. 17, no. 4, pp. 325–336, 2012. DOI: 10.1007/s10666-011-9298-7.

[227] P. Agnolucci, C. Fischer, D. Heine, M. Montes De Oca Leon, J. Pryor, K. Patroni, and S. Hallegatte, "Measuring total carbon pricing," *The World Bank Research Observer*, 2023. DOI: 10.1093/wbro/lkad009.

[228] C. Xiang and T. van Gevelt, "Political signalling and emissions trading schemes in china: Insights from guangdong province," *Energy for Sustainable Development*, vol. 71, pp. 307–314, 2022. DOI: 10.1016/j.esd.2022.10.007.

[229] B. C. Murray and P. T. Maniloff, "Why have greenhouse emissions in RGGI states declined? an econometric attribution to economic, energy market, and policy factors," *Energy Economics*, vol. 51, pp. 581–589, 2015. DOI: 10.1016/j.eneco.2015.07.013.

[230] Y. Zhou and L. Huang, "How regional policies reduce carbon emissions in electricity markets: Fuel switching or emission leakage," *Energy Economics*, vol. 97, p. 105 209, 2021. DOI: 10.1016/j.eneco.2021.105209.

[231] C. Lawley and V. Thivierge, "Refining the evidence: British columbias carbon tax and household gasoline consumption," *The Energy Journal*, vol. 39, no. 2, 2018. DOI: 10.5547/01956574.39.2.claw.

[232] C. Erutku and V. Hildebrand, "Carbon tax at the pump in british columbia and quebec," *Canadian Public Policy*, vol. 44, no. 2, pp. 126–133, 2018. DOI: 10.3138/cpp.2017-027.

[233] R. Best, P. J. Burke, and F. Jotzo, "Carbon pricing efficacy: Cross-country evidence," *Environmental and Resource Economics*, vol. 77, no. 1, pp. 69–94, 2020. DOI: 10.1007/s10640-020-00436-x.

[234] E. Kohlscheen, R. Moessner, and E. Takats, "Effects of carbon pricing and other climate policies on CO 2 emissions," 2021, Publisher: CESifo Working Paper.

[235] P. Runst and A. Thonipara, "Dosis facit effectum why the size of the carbon tax matters: Evidence from the swedish residential sector," *Energy Economics*, vol. 91, p. 104 898, 2020. DOI: 10.1016/j.eneco.2020.104898.

[236] K. P. Gugler, A. Haxhimusa, and M. Liebensteiner, "Carbon pricing and emissions: Causal effects of britain's carbon tax," *SSRN Electronic Journal*, 2022. DOI: 10.2139/ssrn.4116240.

[237] J. Colmer, R. Martin, M. Muûls, and U. J. Wagner, "Does pricing carbon mitigate climate change? firm-level evidence from the european union emissions trading scheme," 2022, Publisher: CEPR Discussion Paper No. DP16982.

[238] Y. Fernández Fernández, M. Fernández López, D. González Hernández, and B. Olmedillas Blanco, "Institutional change and environment: Lessons from the european emission trading system," *Energies*, vol. 11, no. 4, p. 706, 2018. DOI: 10.3390/en11040706.

[239] N. Gupta, J. Shah, S. Gupta, and R. Kaul, "Causal impact of european union emission trading scheme on firm behaviour and economic performance:

A study of german manufacturing firms," 2021, Publisher: arXiv Version Number: 1. DOI: 10.48550/ARXIV.2108.07163.

[240] S. Petrick and U. J. Wagner, "The impact of carbon trading on industry: Evidence from german manufacturing firms," *SSRN Electronic Journal*, 2014. DOI: 10.2139/ssrn.2389800.

[241] K. Vrolijk and M. Sato, "Quasi-experimental evidence on carbon pricing," *The World Bank Research Observer*, vol. 38, no. 2, pp. 213–248, 2023. DOI: 10.1093/wbro/lkad001.

[242] P. J. Ferraro, T. L. Cherry, J. F. Shogren, C. A. Vossler, T. N. Cason, H. B. Flint, J. P. Hochard, O. Johansson-Stenman, P. Martinsson, J. J. Murphy, S. C. Newbold, L. Thunström, D. Van Soest, K. Van 'T Veld, A. Dannenberg, G. F. Loewenstein, and L. Van Boven, "Create a culture of experiments in environmental programs," *Science*, vol. 381, no. 6659, pp. 735–737, 2023. DOI: 10.1126/science.adf7774.

[243] N. Koch, L. Naumann, F. Pretis, N. Ritter, and M. Schwarz, "Attributing agnostically detected large reductions in road CO2 emissions to policy mixes," *Nature Energy*, vol. 7, no. 9, pp. 844–853, 2022. DOI: 10.1038/s41560-022-01095-6.

[244] L. Berrang-Ford, T. Pearce, and J. D. Ford, "Systematic review approaches for climate change adaptation research," *Regional Environmental Change*, vol. 15, no. 5, pp. 755–769, 2015. DOI: 10.1007/s10113-014-0708-7.

[245] A. Sawatzky, A. Cunsolo, A. Jones-Bitton, J. Middleton, and S. L. Harper, "Responding to climate and environmental change impacts on human health via integrated surveillance in the circumpolar north: A systematic realist review," *International Journal of Environmental Research and Public Health*, vol. 15, no. 12, p. 2706, 2018. DOI: 10.3390/ijerph15122706.

[246] W. Muzorewa and M. Chitakira, "Climate-smart livelihood strategies in rural and urban communities in eastern zimbabwe: An in-depth literature study,"

South African Geographical Journal, vol. 103, no. 3, pp. 395–410, 2021. DOI: 10.1080/03736245.2020.1835701.

[247] J. P. Sarmiento Barletti, A. M. Larson, C. Hewlett, and D. Delgado, "Designing for engagement: A realist synthesis review of how context affects the outcomes of multi-stakeholder forums on land use and/or land-use change," *World Development*, vol. 127, p. 104 753, 2020. DOI: 10.1016/j.worlddev.2019.104753.

[248] G. Wong, G. Westhorp, R. Pawson, and T. Greenhalgh, *Realist synthesis: RAMESES training materials*, 2013.

[249] N. Emmel, J. Greenhalgh, A. Manzano, M. Monaghan, and S. Dalkin, "Introduction doing realist evaluation, synthesis and research," in *Doing Realist Research*, N. Emmel, J. Greenhalgh, A. Manzano, M. Monaghan, and S. Dalkin, Eds., 55 City Road: SAGE Publications Ltd, 2018, pp. 1–14. DOI: 10.4135/9781526451729.

[250] G. Wong, "Data gathering in realist reviews: Looking for needles in haystacks," in *Doing Realist Research*, N. Emmel, J. Greenhalgh, A. Manzano, M. Monaghan, and S. Dalkin, Eds., 55 City Road: SAGE Publications Ltd, 2018, pp. 131–146. DOI: 10.4135/9781526451729.

[251] P. Elkins and T. Baker, "Carbon taxes and carbon emissions trading," *Journal of Economic Surveys*, vol. 15, no. 3, pp. 325–376, 2001. DOI: 10.1111/1467-6419.00142.

[252] K. Gugler, A. Haxhimusa, and M. Liebensteiner, "Carbon pricing and emissions: Causal effects of britain's carbon tax," *Energy Economics*, vol. 121, p. 106 655, 2023. DOI: 10.1016/j.eneco.2023.106655.

[253] J. Abrell, M. Kosch, and S. Rausch, "How effective is carbon pricing?—a machine learning approach to policy evaluation," *Journal of Environmental Economics and Management*, vol. 112, p. 102 589, 2022. DOI: 10.1016/j.jeem.2021.102589.

[254] U. J. Wagner, M. Muûls, R. Martin, and J. Colmer, "The causal effect of the european union emissions trading scheme: Evidence from french manufacturing plants," 2014.

[255] J. Colmer, R. Martin, M. Muûls, and U. J. Wagner, "Does pricing carbon mitigate climate change? firm-level evidence from the european union emissions trading scheme," *SSRN Electronic Journal*, 2020. DOI: 10.2139/ssrn.3725482.

[256] N. W. Chan and J. W. Morrow, "Unintended consequences of cap-and-trade? evidence from the regional greenhouse gas initiative," *Energy Economics*, vol. 80, pp. 411–422, 2019. DOI: 10.1016/j.eneco.2019.01.007.

[257] J. Yan, "The impact of climate policy on fossil fuel consumption: Evidence from the regional greenhouse gas initiative (RGGI)," *Energy Economics*, vol. 100, p. 105 333, 2021. DOI: 10.1016/j.eneco.2021.105333.

[258] N. Yajima, T. H. Arimura, and T. Sadayuki, "Energy consumption in transition: Evidence from facility-level data," in *Carbon Pricing in Japan*, T. H. Arimura and S. Matsumoto, Eds., Series Title: Economics, Law, and Institutions in Asia Pacific, Singapore: Springer Singapore, 2021, pp. 129–150. DOI: 10.1007/978-981-15-6964-7_8.

[259] J. J. Andersson, "Carbon taxes and CO2 emissions: Sweden as a case study," *American Economic Journal: Economic Policy*, vol. 11, no. 4, pp. 1–30, 2019. DOI: 10.1257/pol.20170144.

[260] T. K. Mideksa, "Pricing for a cooler planet: An empirical analysis of the effect of taxing carbon," *SSRN Electronic Journal*, 2021. DOI: 10.2139/ssrn.3885415.

[261] Z. Yang, Y. Yuan, and Q. Zhang, "Carbon emission trading scheme, carbon emissions reduction and spatial spillover effects: Quasi-experimental evidence from china," *Frontiers in Environmental Science*, vol. 9, p. 824 298, 2022. DOI: 10.3389/fenvs.2021.824298.

[262] J. Cao, M. S. Ho, R. Ma, and F. Teng, "When carbon emission trading meets a regulated industry: Evidence from the electricity sector of china," *Journal of Public Economics*, vol. 200, p. 104 470, 2021. DOI: 10.1016/j.jpubeco.2021.104470.

[263] W. Xu, "The impact and influencing path of the pilot carbon emission trading market: Evidence from china," *Frontiers in Environmental Science*, vol. 9, p. 787 655, 2021. DOI: 10.3389/fenvs.2021.787655.

[264] C. Wang, Y. Shi, L. Zhang, X. Zhao, and H. Chen, "The policy effects and influence mechanism of china's carbon emissions trading scheme," *Air Quality, Atmosphere & Health*, vol. 14, no. 12, pp. 2101–2114, 2021. DOI: 10.1007/s11869-021-01081-z.

[265] S. Qi, S. Cheng, and J. Cui, "Environmental and economic effects of china's carbon market pilots: Empirical evidence based on a DID model," *Journal of Cleaner Production*, vol. 279, p. 123 720, 2021. DOI: 10.1016/j.jclepro.2020.123720.

[266] Q. Ma, G. Yan, X. Ren, and X. Ren, "Can china's carbon emissions trading scheme achieve a double dividend?" *Environmental Science and Pollution Research*, vol. 29, no. 33, pp. 50 238–50 255, 2022. DOI: 10.1007/s11356-022-19453-y.

[267] Q. Wang, C. Gao, and S. Dai, "Effect of the emissions trading scheme on CO2 abatement in china," *Sustainability*, vol. 11, no. 4, p. 1055, 2019. DOI: 10.3390/su11041055.

[268] Z.-Q. Dong, H. Wang, S.-X. Wang, and L.-H. Wang, "The validity of carbon emission trading policies: Evidence from a quasi-natural experiment in china," *Advances in Climate Change Research*, vol. 11, no. 2, pp. 102–109, 2020. DOI: 10.1016/j.accre.2020.06.001.

[269] J. Cui, C. Wang, J. Zhang, and Y. Zheng, "The effectiveness of china's regional carbon market pilots in reducing firm emissions," *Proceedings of*

the National Academy of Sciences, vol. 118, no. 52, e2109912118, 2021. DOI: 10.1073/pnas.2109912118.

[270] J. Jaraite-Kažukauske and C. Di Maria, "Did the EU ETS make a difference? an empirical assessment using lithuanian firm-level data," *The Energy Journal*, vol. 37, no. 1, 2016. DOI: 10.5547/01956574.37.2.jjar.

[271] M. Klemetsen, K. E. Rosendahl, and A. L. Jakobsen, "The impacts of the EU ETS on norwegian plants' environmental and economic performance," *Climate Change Economics*, vol. 11, no. 1, 2020. DOI: 10.1142/S2010007820500062.

[272] H. Peng, S. Qi, and J. Cui, "The environmental and economic effects of the carbon emissions trading scheme in china: The role of alternative allowance allocation," *Sustainable Production and Consumption*, vol. 28, pp. 105–115, 2021. DOI: 10.1016/j.spc.2021.03.031.

[273] Y. Zhang and J. Zhang, "Estimating the impacts of emissions trading scheme on low-carbon development," *Journal of Cleaner Production*, vol. 238, 2019. DOI: 10.1016/j.jclepro.2019.117913.

[274] W. Zhang, N. Zhang, and Y. Yu, "Carbon mitigation effects and potential cost savings from carbon emissions trading in china's regional industry," *Technological Forecasting and Social Change*, vol. 141, pp. 1–11, 2019. DOI: 10.1016/j.techfore.2018.12.014.

[275] H. Zhang, M. Duan, and Z. Deng, "Have china's pilot emissions trading schemes promoted carbon emission reductions?– the evidence from industrial sub-sectors at the provincial level," *Journal of Cleaner Production*, vol. 234, pp. 912–924, 2019. DOI: 10.1016/j.jclepro.2019.06.247.

[276] H. Zhang, M. Duan, and P. Zhang, "Analysis of the impact of china's emissions trading scheme on reducing carbon emissions," *Energy Procedia*, vol. 158, pp. 3596–3601, 2019. DOI: 10.1016/j.egypro.2019.01.905.

[277] J. Hanoteau and D. Talbot, “Impacts of the québec carbon emissions trading scheme on plant-level performance and employment,” *Carbon Management*, vol. 10, no. 3, pp. 287–298, 2019. DOI: 10.1080/17583004.2019.1595154.

[278] Y. Wang, M. Liao, Y. Wang, L. Xu, and A. Malik, “The impact of foreign direct investment on china’s carbon emissions through energy intensity and emissions trading system,” *Energy Economics*, vol. 97, p. 105 212, 2021. DOI: 10.1016/j.eneco.2021.105212.

[279] K. Tang, Y. Zhou, X. Liang, and D. Zhou, “The effectiveness and heterogeneity of carbon emissions trading scheme in china,” *Environmental Science and Pollution Research*, vol. 28, no. 14, pp. 17 306–17 318, 2021. DOI: 10.1007/s11356-020-12182-0.

[280] T. Abe and T. H. Arimura, “An empirical study of the tokyo emissions trading scheme: An ex post analysis of emissions from university buildings,” in *Carbon Pricing in Japan*, T. H. Arimura and S. Matsumoto, Eds., Singapore: Springer Singapore, 2021, pp. 97–116.

[281] X. Li, Y. Shu, and X. Jin, “Environmental regulation, carbon emissions and green total factor productivity: A case study of china,” *Environment, Development and Sustainability*, vol. 24, no. 2, pp. 2577–2597, 2022. DOI: 10.1007/s10668-021-01546-2.

[282] H. Zhang, R. Zhang, G. Li, W. Li, and Y. Choi, “Has china’s emission trading system achieved the development of a low-carbon economy in high-emission industrial subsectors?” *Sustainability*, vol. 12, no. 13, p. 5370, 2020. DOI: 10.3390/su12135370.

[283] X. Fageda and J. J. Teixidó, “Pricing carbon in the aviation sector: Evidence from the european emissions trading system,” *Journal of Environmental Economics and Management*, vol. 111, p. 102 591, 2022. DOI: 10.1016/j.jeem.2021.102591.

[284] H. Zhang, R. Zhang, G. Li, W. Li, and Y. Choi, "Sustainable feasibility of carbon trading policy on heterogenetic economic and industrial development," *Sustainability*, vol. 11, no. 23, p. 6869, 2019. DOI: 10.3390/su11236869.

[285] G. E. Metcalf, "On the economics of a carbon tax for the united states," *Brookings Papers on Economic Activity*, vol. 2019, no. 1, pp. 405–484, 2019. DOI: 10.1353/eca.2019.0005.

[286] G. E. Metcalf and J. H. Stock, "The macroeconomic impact of europe's carbon taxes," *American Economic Journal: Macroeconomics*, vol. 15, no. 3, pp. 265–286, 2023. DOI: 10.1257/mac.20210052.

[287] S. M. Bartram, K. Hou, and S. Kim, "Real effects of climate policy: Financial constraints and spillovers," *Journal of Financial Economics*, vol. 143, no. 2, pp. 668–696, 2022. DOI: 10.1016/j.jfineco.2021.06.015.

[288] Y. Gao, M. Li, J. Xue, and Y. Liu, "Evaluation of effectiveness of china's carbon emissions trading scheme in carbon mitigation," *Energy Economics*, vol. 90, p. 104 872, 2020. DOI: 10.1016/j.eneco.2020.104872.

[289] T. Sadayuki and T. H. Arimura, "Do regional emission trading schemes lead to carbon leakage within firms? evidence from japan," *Energy Economics*, vol. 104, p. 105 664, 2021. DOI: 10.1016/j.eneco.2021.105664.

[290] S. Clò, M. Ferraris, and M. Florio, "Ownership and environmental regulation: Evidence from the european electricity industry," *Energy Economics*, vol. 61, pp. 298–312, 2017. DOI: 10.1016/j.eneco.2016.12.001.

[291] A. Dechezleprêtre, D. Nachtigall, and F. Venmans, "The joint impact of the european union emissions trading system on carbon emissions and economic performance," *Journal of Environmental Economics and Management*, vol. 118, p. 102 758, 2023. DOI: 10.1016/j.jeem.2022.102758.

[292] F. Pretis, "Does a carbon tax reduce CO2 emissions? evidence from british columbia," *Environmental and Resource Economics*, vol. 83, no. 1, pp. 115–144, 2022. DOI: 10.1007/s10640-022-00679-w.

[293] P. Kim and H. Bae, "Do firms respond differently to the carbon pricing by industrial sector? how and why? a comparison between manufacturing and electricity generation sectors using firm-level panel data in korea," *Energy Policy*, vol. 162, p. 112 773, 2022. DOI: 10.1016/j.enpol.2021.112773.

[294] Q. Wu, K. Tambunlertchai, and P. Pornchaiwiseskul, "Examining the impact and influencing channels of carbon emission trading pilot markets in china," *Sustainability*, vol. 13, no. 10, p. 5664, 2021. DOI: 10.3390/su13105664.

[295] G. Leslie, "Tax induced emissions? estimating short-run emission impacts from carbon taxation under different market structures," *Journal of Public Economics*, vol. 167, pp. 220–239, 2018. DOI: 10.1016/j.jpubeco.2018.09.010.

[296] A. M. Heiaas, "The EU ETS and aviation: Evaluating the effectiveness of the EU emission trading system in reducing emissions from air travel," *Review of Business and Economics Studies*, vol. 9, no. 1, pp. 84–120, 2021. DOI: 10.26794/2308-944X-2021-9-1-84-120.

[297] J. Shen, P. Tang, and H. Zeng, "Does china's carbon emission trading reduce carbon emissions? evidence from listed firms," *Energy for Sustainable Development*, vol. 59, pp. 120–129, 2020. DOI: 10.1016/j.esd.2020.09.007.

[298] X. Yang, P. Jiang, and Y. Pan, "Does china's carbon emission trading policy have an employment double dividend and a porter effect?" *Energy Policy*, vol. 142, p. 111 492, 2020. DOI: 10.1016/j.enpol.2020.111492.

[299] S. Chen, A. Shi, and X. Wang, "Carbon emission curbing effects and influencing mechanisms of china's emission trading scheme: The mediating roles of technique effect, composition effect and allocation effect," *Journal of Cleaner Production*, vol. 264, p. 121 700, 2020. DOI: 10.1016/j.jclepro.2020.121700.

[300] H.-X. Wen, Z.-R. Chen, and P.-Y. Nie, "Environmental and economic performance of china's ETS pilots: New evidence from an expanded synthetic

control method," *Energy Reports*, vol. 7, pp. 2999–3010, 2021. DOI: 10.1016/j.egyr.2021.05.024.

[301] T. H. Arimura and T. Abe, "The impact of the tokyo emissions trading scheme on office buildings: What factor contributed to the emission reduction?" *Environmental Economics and Policy Studies*, vol. 23, no. 3, pp. 517–533, 2021. DOI: 10.1007/s10018-020-00271-w.

[302] P. Bayer and M. Aklin, "The european union emissions trading system reduced CO_2 emissions despite low prices," *Proceedings of the National Academy of Sciences*, vol. 117, no. 16, pp. 8804–8812, 2020. DOI: 10.1073/pnas.1918128117.

[303] B. Hintermann and M. Žarković, "A carbon horse race: Abatement subsidies vs. permit trading in switzerland," *Climate Policy*, vol. 21, no. 3, pp. 290–306, 2021. DOI: 10.1080/14693062.2020.1846485.

[304] M. Ahmad, X. F. Li, and Q. Wu, "Carbon taxes and emission trading systems: Which one is more effective in reducing carbon emissions?—a meta-analysis," *Journal of Cleaner Production*, vol. 476, p. 143 761, 2024. DOI: 10.1016/j.jclepro.2024.143761.

[305] R. Martin, M. Muûls, and U. J. Wagner, "The impact of the european union emissions trading scheme on regulated firms: What is the evidence after ten years?" *Review of Environmental Economics and Policy*, vol. 10, no. 1, pp. 129–148, 2016. DOI: 10.1093/reep/rev016.

[306] T. Laing, M. Sato, M. Grubb, and C. Comberti, "The effects and side-effects of the EU emissions trading scheme," *WIREs Climate Change*, vol. 5, no. 4, pp. 509–519, 2014. DOI: 10.1002/wcc.283.

[307] J. Arlinghaus, "Impacts of carbon prices on indicators of competitiveness: A review of empirical findings," OECD Environment Working Papers 87, 2015, Series: OECD Environment Working Papers Volume: 87. DOI: 10.1787/5js37p21grzq-en.

[308] F. Venmans, "A literature-based multi-criteria evaluation of the EU ETS," *Renewable and Sustainable Energy Reviews*, vol. 16, no. 8, pp. 5493–5510, 2012. DOI: 10.1016/j.rser.2012.05.036.

[309] J. van den Bergh, J. Castro, S. Drews, F. Exadaktylos, J. Foramitti, F. Klein, T. Konc, and I. Savin, "Designing an effective climate-policy mix: Accounting for instrument synergy," *Climate Policy*, vol. 21, no. 6, pp. 745–764, 2021. DOI: 10.1080/14693062.2021.1907276.

[310] P. Del Río González, "The interaction between emissions trading and renewable electricity support schemes. an overview of the literature," *Mitigation and Adaptation Strategies for Global Change*, vol. 12, no. 8, pp. 1363–1390, 2007. DOI: 10.1007/s11027-006-9069-y.

[311] R. C. Repetto and D. Austin, *The costs of climate protection: a guide for the perplexed*. Washington, DC: World Resources Institute, 1997, 51 pp.

[312] C. Fischer and R. D. Morgenstern, "Carbon abatement costs: Why the wide range of estimates?" *The Energy Journal*, vol. 27, no. 2, pp. 73–86, 2006. DOI: 10.5547/ISSN0195-6574-EJ-Vol27-No2-5.

[313] S. K. Huang, L. Kuo, and K.-L. Chou, "The applicability of marginal abatement cost approach: A comprehensive review," *Journal of Cleaner Production*, vol. 127, pp. 59–71, 2016. DOI: 10.1016/j.jclepro.2016.04.013.

[314] G. Wong, T. Greenhalgh, G. Westhorp, J. Buckingham, and R. Pawson, "RAMESES publication standards: Realist syntheses," *BMC Medicine*, vol. 11, no. 1, p. 21, 2013. DOI: 10.1186/1741-7015-11-21.

[315] S. Nutley, I. Walter, and H. T. O. Davies, "From knowing to doing: A framework for understanding the evidence-into-practice agenda," *Evaluation*, vol. 9, no. 2, pp. 125–148, 2003. DOI: 10.1177/1356389003009002002.

[316] K. Oliver, S. Innvar, T. Lorenc, J. Woodman, and J. Thomas, "A systematic review of barriers to and facilitators of the use of evidence by policymakers,"

BMC Health Services Research, vol. 14, no. 1, p. 2, 2014. DOI: 10.1186/1472-6963-14-2.

[317] N. R. Haddaway and A. S. Pullin, "The policy role of systematic reviews: Past, present and future," *Springer Science Reviews*, vol. 2, no. 1, pp. 179–183, 2014. DOI: 10.1007/s40362-014-0023-1.

[318] D. Huitema, A. Jordan, E. Massey, T. Rayner, H. v. Asselt, C. Haug, R. Hildingsson, S. Monni, and J. Stripple, "The evaluation of climate policy: Theory and emerging practice in europe," *Policy Sciences*, vol. 44, no. 2, pp. 179–198, 2011. DOI: 10.1007/s11077-011-9125-7.

[319] M. Callaghan, C.-F. Schleussner, S. Nath, Q. Lejeune, T. R. Knutson, M. Reichstein, G. Hansen, E. Theokritoff, M. Andrijevic, R. J. Brecha, M. Hegarty, C. Jones, K. Lee, A. Lucas, N. Van Maanen, I. Menke, P. Pfleiderer, B. Yesil, and J. C. Minx, "Machine-learning-based evidence and attribution mapping of 100,000 climate impact studies," *Nature Climate Change*, vol. 11, no. 11, pp. 966–972, 2021. DOI: 10.1038/s41558-021-01168-6.

[320] G. C. Nunez–Mir, B. V. Iannone, B. C. Pijanowski, N. Kong, and S. Fei, "Automated content analysis: Addressing the big literature challenge in ecology and evolution," *Methods in Ecology and Evolution*, vol. 7, no. 11, pp. 1262–1272, 2016. DOI: 10.1111/2041-210X.12602.

[321] K. B. Cohen, H. L. Johnson, K. Verspoor, C. Roeder, and L. E. Hunter, "The structural and content aspects of abstracts versus bodies of full text journal articles are different," *BMC Bioinformatics*, vol. 11, no. 1, p. 492, 2010. DOI: 10.1186/1471-2105-11-492.

[322] C. Blake, "Beyond genes, proteins, and abstracts: Identifying scientific claims from full-text biomedical articles," *Journal of Biomedical Informatics*, vol. 43, no. 2, pp. 173–189, 2010. DOI: 10.1016/j.jbi.2009.11.001.

[323] H. Chen, M. Di Gregorio, and J. Paavola, "Carbon leakage in emissions trading systems: A systematic review and meta-analysis of ex-ante and ex-post

evidence," [manuscript submitted for publication], School of Earth and Environment, University of Leeds, 2024.

[324] Z. Chen, P. E. Brockway, S. Few, and J. Paavola, "The impact of emissions trading systems on technological innovation for climate change mitigation: A systematic review," [manuscript submitted for publication], School of Earth and Environment, University of Leeds, 2024.

[325] E. Narassimhan, K. S. Gallagher, S. Koester, and J. R. Alejo, "Carbon pricing in practice: A review of existing emissions trading systems," *Climate Policy*, vol. 18, no. 8, pp. 967–991, 2018. DOI: 10.1080/14693062.2018.1467827.

[326] I. Savin, S. Drews, and J. Van Den Bergh, "Carbon pricing – perceived strengths, weaknesses and knowledge gaps according to a global expert survey," *Environmental Research Letters*, vol. 19, no. 2, p. 024 014, 2024. DOI: 10.1088/1748-9326/ad1c1c.

[327] M. J. Fell, "Anticipating distributional impacts of peer-to-peer energy trading: Inference from a realist review of evidence on airbnb," *Cleaner and Responsible Consumption*, vol. 2, p. 100 013, 2021. DOI: 10.1016/j.clrc.2021.100013.

[328] P. O'Campo, A. Molnar, E. Ng, E. Renahy, C. Mitchell, K. Shankardass, A. St. John, C. Bambra, and C. Muntaner, "Social welfare matters: A realist review of when, how, and why unemployment insurance impacts poverty and health," *Social Science & Medicine*, vol. 132, pp. 88–94, 2015. DOI: 10.1016/j.socscimed.2015.03.025.

[329] R. McLain, S. Lawry, and M. Ojanen, "Fisheries' property regimes and environmental outcomes: A realist synthesis review," *World Development*, vol. 102, pp. 213–227, 2018. DOI: 10.1016/j.worlddev.2017.09.016.

[330] R. C. Berg and E. Denison, "Interventions to reduce the prevalence of female genital mutilation/cutting in african countries," *Campbell Systematic Reviews*, vol. 8, no. 1, pp. 1–155, 2012. DOI: 10.4073/csr.2012.9.

[331] M. J. Fell, K. Roelich, and L. Middlemiss, "Realist approaches in energy research to support faster and fairer climate action," *Nature Energy*, vol. 7, no. 10, pp. 916–922, 2022. DOI: 10.1038/s41560-022-01093-8.

[332] P. Warren, "The use of systematic reviews to analyse demand-side management policy," *Energy Efficiency*, vol. 7, no. 3, pp. 417–427, 2014. DOI: 10.1007/s12053-013-9230-x.

[333] N. Emmel, J. Greenhalgh, A. Manzano, M. Monaghan, and S. Dalkin, Eds., *Doing realist research*, First edition, Los Angeles: Sage, 2018, 251 pp.

[334] J. Porciello, M. Ivanina, M. Islam, S. Einarson, and H. Hirsh, "Accelerating evidence-informed decision-making for the sustainable development goals using machine learning," *Nature Machine Intelligence*, vol. 2, no. 10, pp. 559–565, 2020. DOI: 10.1038/s42256-020-00235-5.

[335] G. Tsafnat, P. Glasziou, M. K. Choong, A. Dunn, F. Galgani, and E. Coiera, "Systematic review automation technologies," *Systematic reviews*, vol. 3, no. 1, p. 74, 2014.

[336] W. X. Zhao, K. Zhou, J. Li, T. Tang, X. Wang, Y. Hou, Y. Min, B. Zhang, J. Zhang, Z. Dong, Y. Du, C. Yang, Y. Chen, Z. Chen, J. Jiang, R. Ren, Y. Li, X. Tang, Z. Liu, P. Liu, J.-Y. Nie, and J.-R. Wen, *A survey of large language models*, Version Number: 15, 2023. DOI: 10.48550/ARXIV.2303.18223.

[337] O. Baclic, M. Tunis, K. Young, C. Doan, and H. Swerdfeger, "Challenges and opportunities for public health made possible by advances in natural language processing," *Canada Communicable Disease Report*, pp. 161–168, 2020. DOI: 10.14745/ccdr.v46i06a02.

[338] E. Beller, J. Clark, G. Tsafnat, C. Adams, H. Diehl, H. Lund, M. Ouzzani, K. Thayer, J. Thomas, T. Turner, J. Xia, K. Robinson, and P. Glasziou, "Making progress with the automation of systematic reviews: Principles of the international collaboration for the automation of systematic reviews (ICASR),"

Systematic Reviews, vol. 7, no. 1, p. 77, 2018. DOI: 10.1186/s13643-018-0740-7.

[339] S. Chabou and M. Iglewski, "Combination of conditional random field with a rule based method in the extraction of PICO elements," *BMC Medical Informatics and Decision Making*, vol. 18, no. 1, p. 128, 2018. DOI: 10.1186/s12911-018-0699-2.

[340] D. Jin and P. Szolovits, "Advancing PICO element detection in biomedical text via deep neural networks," *Bioinformatics*, vol. 36, no. 12, Z. Lu, Ed., pp. 3856–3862, 2020. DOI: 10.1093/bioinformatics/btaa256.

[341] J. Thomas, J. McNaught, and S. Ananiadou, "Applications of text mining within systematic reviews," *Research Synthesis Methods*, vol. 2, no. 1, pp. 1–14, 2011. DOI: 10.1002/jrsm.27.

[342] I. J. Marshall, A. Noel-Storr, J. Kuiper, J. Thomas, and B. C. Wallace, "Machine learning for identifying randomized controlled trials: An evaluation and practitioner's guide," *Research Synthesis Methods*, vol. 9, no. 4, pp. 602–614, 2018. DOI: 10.1002/jrsm.1287.

[343] M. Stevenson and R. Bin-Hezam, "Stopping methods for technology-assisted reviews based on point processes," *ACM Transactions on Information Systems*, vol. 42, no. 3, pp. 1–37, 2024. DOI: 10.1145/3631990.

[344] A. Sneyd and M. Stevenson, "Modelling stopping criteria for search results using poisson processes," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, Hong Kong, China: Association for Computational Linguistics, 2019, pp. 3482–3487. DOI: 10.18653/v1/D19-1351.

[345] D. D. Lewis, L. Gray, and M. Noel, "Confidence sequences for evaluating one-phase technology-assisted review," in *Proceedings of the Nineteenth In-*

ternational Conference on Artificial Intelligence and Law*, Braga Portugal: ACM, 2023, pp. 131–140. DOI: 10.1145/3594536.3595167.

[346] S. Sharpe, L. Benton, and A. Haines, "Evidence must guide policy and practice towards health centred and equitable climate solutions," *BMJ*, q1417, 2024. DOI: 10.1136/bmj.q1417.

[347] C. Lefebvre, J. Glanville, S. Briscoe, R. Featherstone, A. Littlewood, M.-I. Metzendorf, A. Noel-Storr, R. Paynter, T. Rader, J. Thomas, and L. S. Wieland, "Chapter 4: Searching for and selecting studies," in *Cochrane Handbook for Systematic Reviews of Interventions*, J. P. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, and V. A. Welch, Eds., 6.5, 2024.

[348] K. Dickersin, R. Scherer, and C. Lefebvre, "Systematic reviews: Identifying relevant studies for systematic reviews," *BMJ*, vol. 309, no. 6964, pp. 1286–1291, 1994. DOI: 10.1136/bmj.309.6964.1286.

[349] M. Maier, F. Bartoš, and E.-J. Wagenmakers, "Robust bayesian meta-analysis: Addressing publication bias with model-averaging.," *Psychological Methods*, vol. 28, no. 1, pp. 107–122, 2023. DOI: 10.1037/met0000405.

[350] J. P. Higgins, J. Savović, M. J. Page, R. G. Elbers, and J. A. Sterne, "Assessing risk of bias in a randomized trial," in *Cochrane Handbook for Systematic Reviews of Interventions*, J. P. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, and V. A. Welch, Eds., 6.5, [last updated October 2019], Wiley, 2024.

[351] T. H. Barker, J. C. Stone, K. Sears, M. Klugar, J. Leonardi-Bee, C. Tufanaru, E. Aromataris, and Z. Munn, "Revising the JBI quantitative critical appraisal tools to improve their applicability: An overview of methods and the development process," *JBI Evidence Synthesis*, vol. 21, no. 3, pp. 478–493, 2023. DOI: 10.11124/JBIES-22-00125.

[352] E. E. Leamer, "Let's take the con out of econometrics," *The American Economic Review*, vol. 73, no. 1, pp. 31–43, 1983.

[353] S. Athey and G. W. Imbens, "The state of applied econometrics: Causality and policy evaluation," *Journal of Economic Perspectives*, vol. 31, no. 2, pp. 3–32, 2017. DOI: 10.1257/jep.31.2.3.

[354] X. Sala-i-Martin, "I just ran four million regressions," National Bureau of Economic Research, Cambridge, MA, w6252, 1997, w6252. DOI: 10.3386/w6252.

[355] R. Levine and D. Renelt, "A sensitivity analysis of cross-country growth regressions," *The American Economic Review*, vol. 82, no. 4, pp. 942–963, 1992.

[356] S. N. Durlauf, P. A. Johnson, and J. R. Temple, "Chapter 8 growth econometrics," in *Handbook of Economic Growth*, vol. 1, Elsevier, 2005, pp. 555–677. DOI: 10.1016/S1574-0684(05)01008-7.

[357] S. B. Bruns and J. P. Ioannidis, "Determinants of economic growth: Different time different answer?" *Journal of Macroeconomics*, vol. 63, p. 103 185, 2020. DOI: 10.1016/j.jmacro.2019.103185.

[358] S. Dhakal, J. C. Minx, F. L. Toth, A. Abdel-Aziz, M. J. Figueroa Meza, K. Hubacek, I. G. Jonckheere, Y.-G. Kim, G. F. Nemet, S. Pachauri, X. C. Tan, and T. Wiedmann, "Emissions trends and drivers," in *Climate Change 2022 - Mitigation of Climate Change*, IPCC, Ed., 1st ed., Cambridge University Press, 2023, pp. 215–294. DOI: 10.1017/9781009157926.004.

[359] K. Dong, X. Dong, and C. Dong, "Determinants of the global and regional CO_2 emissions: What causes what and where?" *Applied Economics*, vol. 51, no. 46, pp. 5031–5044, 2019. DOI: 10.1080/00036846.2019.1606410.

[360] Q. Xia, H. Wang, X. Liu, and X. Pan, "Drivers of global and national CO_2 emissions changes 2000–2017," *Climate Policy*, vol. 21, no. 5, pp. 604–615, 2021. DOI: 10.1080/14693062.2020.1864267.

[361] A. Malik and J. Lan, "The role of outsourcing in driving global carbon emissions," *Economic Systems Research*, vol. 28, no. 2, pp. 168–182, 2016. DOI: 10.1080/09535314.2016.1172475.

[362] E. A. Rosa and T. Dietz, "Human drivers of national greenhouse-gas emissions," *Nature Climate Change*, vol. 2, no. 8, pp. 581–586, 2012. DOI: 10.1038/nclimate1506.

[363] D. Card, "Minimum wages and employment: A case study of the fast food industry in new jersey and pennsylvania," 1994.

[364] A. Abadie, A. Diamond, and J. Hainmueller, "Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program," *Journal of the American Statistical Association*, vol. 105, no. 490, pp. 493–505, 2010. DOI: 10.1198/jasa.2009.ap08746.

[365] G. W. Imbens and J. M. Wooldridge, "Recent developments in the econometrics of program evaluation," *Journal of Economic Literature*, vol. 47, no. 1, pp. 5–86, 2009. DOI: 10.1257/jel.47.1.5.

[366] G. W. Imbens and T. Lemieux, "Regression discontinuity designs: A guide to practice," *Journal of Econometrics*, vol. 142, no. 2, pp. 615–635, 2008. DOI: 10.1016/j.jeconom.2007.05.001.

[367] B. D. Meyer, "Natural and quasi-experiments in economics," *Journal of Business & Economic Statistics*, vol. 13, no. 2, pp. 151–161, 1995. DOI: 10.1080/07350015.1995.10524589.

[368] G. W. Imbens and D. B. Rubin, *Causal inference in statistics, social, and biomedical sciences*. Cambridge university press, 2015.

[369] D. J. Cook, C. D. Mulrow, and R. B. Haynes, "Systematic reviews: Synthesis of best evidence for clinical decisions," *Annals of Internal Medicine*, vol. 126, no. 5, pp. 376–380, 1997. DOI: 10.7326/0003-4819-126-5-199703010-00006.

[370] D. Gough, J. Thomas, and S. Oliver, "An introduction to systematic reviews," 2017, Publisher: SAGE Publications ltd.

[371] T. D. Stanley, *Meta-regression analysis in economics and business*, in collab. with H. Doucouliagos. London New York: Routledge, 2012, 1 p. DOI: 10.4324/9780203111710.

[372] J. D. Ford, L. Cameron, J. Rubis, M. Maillet, D. Nakashima, A. C. Willox, and T. Pearce, "Including indigenous knowledge and experience in IPCC assessment reports," *Nature Climate Change*, vol. 6, no. 4, pp. 349–353, 2016. DOI: 10.1038/nclimate2954.

[373] J. Petzold, N. Andrews, J. D. Ford, C. Hedemann, and J. C. Postigo, "Indigenous knowledge on climate change adaptation: A global evidence map of academic literature," *Environmental Research Letters*, vol. 15, no. 11, p. 113 007, 2020. DOI: 10.1088/1748-9326/abb330.

[374] P. Rashidi and K. Lyons, "Democratizing global climate governance? the case of indigenous representation in the intergovernmental panel on climate change (IPCC)," in *Economics and Climate Emergency*, 1st ed. London: Routledge, 2022, pp. 284–299. DOI: 10.4324/9781003174707-21.

[375] G. Martin and E. Saikawa, "Effectiveness of state climate and energy policies in reducing power-sector CO2 emissions," *Nature Climate Change*, vol. 7, no. 12, pp. 912–919, 2017. DOI: 10.1038/s41558-017-0001-0.

[376] F. Dong, Y. Dai, S. Zhang, X. Zhang, and R. Long, "Can a carbon emission trading scheme generate the porter effect? evidence from pilot areas in china," *Science of The Total Environment*, vol. 653, pp. 565–577, 2019. DOI: 10.1016/j.scitotenv.2018.10.395.

[377] L. Zhao and C. Yang, "Research on fossil fuel related carbon emissions reduction scheme effects," *China Petroleum Processing & Petrochemical Technology*, vol. 22, no. 1, pp. 32–39, 2020.

[378] X. Ouyang, X. Fang, Y. Cao, and C. Sun, "Factors behind CO2 emission reduction in chinese heavy industries: Do environmental regulations matter?" *Energy Policy*, vol. 145, p. 111 765, 2020. DOI: 10.1016/j.enpol.2020.111765.

[379] X. Wang, J. Huang, and H. Liu, "Can china's carbon trading policy help achieve carbon neutrality? — a study of policy effects from the five-sphere integrated plan perspective," *Journal of Environmental Management*, vol. 305, p. 114 357, 2022. DOI: 10.1016/j.jenvman.2021.114357.

[380] Y. Wen, P. Hu, J. Li, Q. Liu, L. Shi, J. Ewing, and Z. Ma, "Does china's carbon emissions trading scheme really work? a case study of the hubei pilot," *Journal of Cleaner Production*, vol. 277, p. 124 151, 2020. DOI: 10.1016/j.jclepro.2020.124151.

[381] X. Yang, P. Jiang, and Y. Pan, "Does china's carbon emission trading policy have an employment double dividend and a porter effect?" *Energy Policy*, vol. 142, p. 111 492, 2020. DOI: 10.1016/j.enpol.2020.111492.

[382] B. Yang, L. Liu, and Y. Yin, "Will china's low-carbon policy balance emission reduction and economic development? evidence from two provinces," *International Journal of Climate Change Strategies and Management*, vol. 13, no. 1, pp. 78–94, 2021. DOI: 10.1108/IJCCSM-08-2020-0093.

[383] L. Yi, N. Bai, L. Yang, Z. Li, and F. Wang, "Evaluation on the effectiveness of china's pilot carbon market policy," *Journal of Cleaner Production*, vol. 246, p. 119 039, 2020. DOI: 10.1016/j.jclepro.2019.119039.

[384] W. Zhang, J. Li, G. Li, and S. Guo, "Emission reduction effect and carbon market efficiency of carbon emissions trading policy in china," *Energy*, vol. 196, p. 117 117, 2020. DOI: 10.1016/j.energy.2020.117117.

[385] Y. Zhang, S. Li, T. Luo, and J. Gao, "The effect of emission trading policy on carbon emission reduction: Evidence from an integrated study of pilot

regions in china," *Journal of Cleaner Production*, vol. 265, p. 121 843, 2020. DOI: 10.1016/j.jclepro.2020.121843.

[386] I. Ko and T. Lee, "Carbon pricing and decoupling between greenhouse gas emissions and economic growth: A panel study of 29 european countries, 1996–2014," *Review of Policy Research*, vol. 39, no. 5, pp. 654–673, 2022. DOI: 10.1111/ropr.12458.

[387] G. Metcalf and J. Stock, "The macroeconomic impact of europe's carbon taxes," National Bureau of Economic Research, 2020. DOI: 10.3386/w27488.

[388] J.-D. Elbaum, "The effect of a carbon tax on per capita carbon dioxide emissions: Evidence from finland," University of Neuchâtel, Institute of Economic Research (IRENE), Neuchâtel, IRENE Working Paper 21-05, 2021.

[389] M. Hamamoto, "Target-setting emissions trading program in saitama prefecture: Impact on CO2 emissions in the first compliance period," in *Carbon Pricing in Japan*, T. H. Arimura and S. Matsumoto, Eds., Series Title: Economics, Law, and Institutions in Asia Pacific, Singapore: Springer Singapore, 2021, pp. 117–127. DOI: 10.1007/978-981-15-6964-7_7.

Appendix

Appendix for Chapter 2

Group	Schemes	Keywords
Argentina	Argentina carbon tax	Argentina
Australia	Australia ETS	Australia
Austria	Austria ETS	Austria
Canada	Alberta ETS, British Columbia carbon tax, Canada federal carbon tax, Canada federal ETS, New Brunswick carbon tax, New Brunswick ETS, Newfoundland and Labrador carbon tax, Newfoundland and Labrador ETS, Northwest Territories carbon tax, Nova Scotia ETS, Ontario ETS, Prince Edward Island carbon tax, Quebec ETS, Saskatchewan ETS	Quebec, Alberta, Nova Scotia, Prince Edward Island, Northwest, Newfoundland, Canada, New Brunswick, Ontario, Saskatchewan, British Columbia, Canadian
Chile	Chile carbon tax	Chile
China	Beijing pilot ETS, China national ETS, China regional ETS pilots, Chongqing pilot ETS, Fujian pilot ETS, Guangdong pilot ETS, Hubei pilot ETS, Shanghai pilot ETS, Shenzhen pilot ETS, Tianjin pilot ETS	Hubei, Shenzhen, Guangdong, Fujian, Beijing, China, Chongqing, Tianjin, Shanghai, Chinese
Colombia	Colombia carbon tax	Colombia
Denmark	Denmark carbon tax	Denmark, Danish
Estonia	Estonia carbon tax	Estonia
EU	EU ETS	Europe, EU, EUA
Finland	Finland carbon tax	Finland, Finnish
France	France carbon tax	France, French
Germany	Germany ETS	Germany, German
Iceland	Iceland carbon tax	Iceland
Indonesia	Indonesia ETS	Indonesia
Ireland	Ireland carbon tax	Ireland, Irish
Japan	Saitama ETS, Japan carbon tax	Japan, Saitama
Kazakhstan	Kazakhstan ETS	Kazakhstan
Korea	Korea ETS	Korea
Latvia	Latvia carbon tax	Latvia
Liechtenstein	Liechtenstein carbon tax	Liechtenstein
Luxembourg	Luxembourg carbon tax	Luxembourg
Mexico	Mexico carbon tax, Mexico ETS, Baja California carbon tax, Tamaulipas carbon tax, Zacatecas carbon tax	Mexico, Baja, California, Mexican
Netherlands	Netherlands carbon tax	Netherlands, Dutch
New Zealand	New Zealand ETS	New Zealand
Norway	Norway carbon tax	Norway, Norwegian
Poland	Poland carbon tax	Poland, Polish
Portugal	Portugal carbon tax	Portugal, Portuguese
Singapore	Singapore carbon tax	Singapore
Slovenia	Slovenia carbon tax	Slovenia
South Africa	South Africa carbon tax	South Africa
Spain	Spain carbon tax	Spain, Spanish
Sweden	Sweden carbon tax	Sweden, Swedish
Switzerland	Switzerland carbon tax, Switzerland ETS	Switzerland
Tokyo	Tokyo ETS	Tokyo
UK	UK carbon price support, UK ETS	United Kingdom, England, Wales
Ukraine	Ukraine carbon tax	Ukraine
unknown	multiple, unclear	
Uruguay	Uruguay carbon tax	Uruguay
US	California ETS, RGGI, Massachusetts ETS, Oregon ETS	Massachusetts, New York, RGGI, Maryland, California, Regional Greenhouse Gas, Greenhouse Gas Initiative

Table A.1: **Keywords for dictionary approach to identify policy regions for evidence and gap map**

IPCC	Penasco et al	Konidari & Mavrakis		Schneider & Wagner	Neij & Åstrand	Köppl & Schratzenstaller	Lamb et al.	Synthesised Typology
Category	Category	Category	Sub-Category	Category	Category	Category	Category	Category
Environmental effectiveness	Environmental effectiveness	Environmental performance	Direct contribution to reduction of GHG emissions	Environmental efficacy	Relevance	Environmental effectiveness		Environmental effectiveness
			Indirect environmental effects of the instrument		Sustainability			
								Leakage
Economic effectiveness	Competitiveness	Political acceptability	Competitiveness			Effects of carbon taxes on macroeconomic performance		Competitiveness
						The double dividend hypothesis		
						Effects of carbon taxes on competitiveness		
	Cost related outcomes		Cost effectiveness	Cost effectiveness	Efficiency			Cost effectiveness & Efficiency
			Dynamic cost efficiency					
							Employment	Employment & Labour market
Distributional effects	Distributional outcomes		Equity			Distributional effects of carbon taxation	Equality (income)	Distributional effects & fairness
						Distributional consequences of compen-sation measures	Equality (geographic)	
							Equality (gender)	
	Other social outcomes						Access (electricity)	
							Access (other)	
							Affordability (energy)	
							Time/labour/ drudgery	
							Subjective well-being	
							Livelihoods and poverty	
							Community cohesion/conflict	
Transformative potential	Innovation outcomes					Effects of carbon taxes on innovation		Innovation & Investment
	Technological effectiveness			Dynamic efficiency	Market indicators			Firm behaviour & economic structure
					Business oriented indicators			
								Prices of goods and services
								Household behaviour
Institutional requirements				Political economy aspects	implementation process	acceptance and public support of carbon tax schemes		Implementation process & Feasibility
			Flexibility					
			Stringency for non-compliance and non-participation					
		Feasibility of implementation (or enforcement)	Implementation network capacity	Administrative practicability				
			Administrative feasibility					
			Financial feasibility					
							Procedural justice	
								(Public) Perception
Co-benefits, negative side-effects								Environmental and health co-benefits
								other

Figure A.1: **Synthesis of outcome categories for evidence and gap map**: The overview lists outcome typologies proposed by Dubash *et al.* [9] (IPCC), Peñasco *et al.* [140], Konidari and Mavrakis [150], Schneider and Wagner [153], Neij and Åstrand [152], Köppl and Schratzenstaller [138], and Lamb *et al.* [151]. A synthesis of these typologies, used for the evidence and gap map, is provided in the right column.

Appendix for Chapter 3

Table B.1: **Studies included in the systematic reviews in Chapter 3 and 4**

Study	Scheme	Studied sector	Study design
Leslie [295]	Australian carbon tax	Energy	control/treatment
Lawley and Thivierge [231]	BC carbon tax	Transport	price elasticity
Metcalf [285]	BC carbon tax	Whole economy	DiD
Pretis [292]	BC carbon tax	All sectors covered by the carbon price	DiD
Rivers and Schaufele [176]	BC carbon tax	Transport	price elasticity
Xiang and Lawley [177]	BC carbon tax	Buildings	DiD, price elasticity
Erutku and Hilde-brand [232]	BC carbon tax, Quebec ETS	Transport	price elasticity
Bartram *et al.* [287]	California CaT	Industry	DiD
Martin and Saikawa [375]	California CaT, RGGI	Energy	DiD
Cao *et al.* [262]	Chinese pilot ETS	Energy	DiD
Chen *et al.* [299]	Chinese pilot ETS	Whole economy	DiD
Cui *et al.* [269]	Chinese pilot ETS	All sectors covered by the carbon price	DiD
Dong *et al.* [376]	Chinese pilot ETS	Whole economy	DiD
Dong *et al.* [268]	Chinese pilot ETS	All sectors covered by the carbon price	DiD
Gao *et al.* [288]	Chinese pilot ETS	All sectors covered by the carbon price	DiD
Hu *et al.* [174]	Chinese pilot ETS	All sectors covered by the carbon price	DiD
Li *et al.* [281]	Chinese pilot ETS	All sectors covered by the carbon price	DiD
Zhao and Yang [377]	Chinese pilot ETS	All sectors covered by the carbon price	DiD
Ma *et al.* [266]	Chinese pilot ETS	All sectors covered by the carbon price	DiD
Ouyang *et al.* [378]	Chinese pilot ETS	Industry	DiD

Table B.1 – *Continued from previous page*

Study	Scheme	Studied sector	Study design
Peng *et al.* [272]	Chinese pilot ETS	All sectors covered by the carbon price	DiD
Qi *et al.* [265]	Chinese pilot ETS	Whole economy	DiD
Shen *et al.* [297]	Chinese pilot ETS	All sectors covered by the carbon price	DiD
Tang *et al.* [279]	Chinese pilot ETS	All sectors covered by the carbon price	DiD
Wang *et al.* [267]	Chinese pilot ETS	Whole economy	DiD
Wang *et al.* [379]	Chinese pilot ETS	Whole economy	DiD
Wang *et al.* [278]	Chinese pilot ETS	Whole economy	DiD
Wang *et al.* [264]	Chinese pilot ETS	All sectors covered by the carbon price	DiD
Wen *et al.* [380]	Chinese pilot ETS	Industry	DiD
Wen *et al.* [300]	Chinese pilot ETS	Industry	DiD
Wu *et al.* [294]	Chinese pilot ETS	All sectors covered by the carbon price	DiD
Xu [263]	Chinese pilot ETS	All sectors covered by the carbon price	DiD
Yang *et al.* [381]	Chinese pilot ETS	All sectors covered by the carbon price	DiD
Yang *et al.* [382]	Chinese pilot ETS	All sectors covered by the carbon price	DiD
Yang *et al.* [261]	Chinese pilot ETS	All sectors covered by the carbon price	DiD
Yi *et al.* [383]	Chinese pilot ETS	All sectors covered by the carbon price	DiD
Zhang and Zhang [273]	Chinese pilot ETS	Whole economy	DiD
Zhang *et al.* [275]	Chinese pilot ETS	All sectors covered by the carbon price	DiD
Zhang *et al.* [276]	Chinese pilot ETS	All sectors covered by the carbon price	DiD
Zhang *et al.* [384]	Chinese pilot ETS	Industry	DiD
Zhang *et al.* [385]	Chinese pilot ETS	Whole economy	DiD
Zhang *et al.* [274]	Chinese pilot ETS	Industry	DiD
Zhang *et al.* [284]	Chinese pilot ETS	Whole economy	DiD

Table B.1 – *Continued from previous page*

Study	Scheme	Studied sector	Study design
Zhang *et al.* [282]	Chinese pilot ETS	All sectors covered by the carbon price	DiD
Best *et al.* [233]	cross-country	Whole economy	control/treatment
Ko and Lee [386]	cross-country	Whole economy	control/treatment
Metcalf and Stock [387]	cross-country	All sectors covered by the carbon price	before/after treatment
Rafaty *et al.* [18]	cross-country	All sectors covered by the carbon price	DiD
Bayer and Aklin [302]	EU ETS	All sectors covered by the carbon price	DiD
Clò *et al.* [290]	EU ETS	Energy	before/after treatment
Colmer *et al.* [237]	EU ETS	Industry	DiD
Dechezleprêtre *et al.* [291]	EU ETS	All sectors covered by the carbon price	DiD
Fageda and Teixidó [283]	EU ETS	Transport	DiD
Fernández Fernández *et al.* [238]	EU ETS	Whole economy	before/after treatment
Gupta *et al.* [239]	EU ETS	Industry	DiD
Heiaas [296]	EU ETS	Transport	DiD
Jaraite-Kažukauske and Di Maria [270]	EU ETS	All sectors covered by the carbon price	DiD
Klemetsen *et al.* [271]	EU ETS	All sectors covered by the carbon price	DiD
Petrick and Wagner [240]	EU ETS	Industry	DiD
Wagner *et al.* [254]	EU ETS	Industry	DiD
Gugler *et al.* [178]	EU ETS, UK carbon price support	Energy	price elasticity
Elbaum [388]	Finnish carbon tax	Transport	DiD
Mideksa [260]	Finnish carbon tax	Transport	DiD
Kim and Bae [293]	Korea ETS	Industry	
Hanoteau and Talbot [277]	Quebec ETS	All sectors covered by the carbon price	DiD
Chan and Morrow [256]	RGGI	All sectors covered by the carbon price	DiD

Table B.1 – *Continued from previous page*

Study	Scheme	Studied sector	Study design
Murray and Maniloff [229]	RGGI	Energy	DiD
Yan [257]	RGGI	Energy	DiD
Zhou and Huang [230]	RGGI	Energy	DiD
Hamamoto [389]	Saitama ETS	All sectors covered by the carbon price	DiD
Sadayuki and Arimura [289]	Saitama ETS, Tokyo ETS	All sectors covered by the carbon price	DiD
Yajima *et al.* [258]	Saitama ETS, Tokyo ETS	Industry	DiD
Andersson [259]	Swedish carbon tax	Transport	DiD
Runst and Thonipara [235]	Swedish carbon tax	Buildings	DiD
Hintermann and Žarković [303]	Swiss ETS	All sectors covered by the carbon price	DiD
Abe and Arimura [280]	Tokyo ETS	Buildings	DiD
Arimura and Abe [301]	Tokyo ETS	Industry	DiD
Abrell *et al.* [253]	UK carbon price support	Energy	before/after treatment
Gugler *et al.* [236]	UK carbon price support	Energy	before/after treatment
Leroutier [175]	UK carbon price support	Energy	DiD

	Full sample	Full sample with outliers	Full sample clustered	Full sample, PET	Full sample, PEESE	Full sample, fixed effects (clustered)	Sub-sample: low RoB	Sub-sample: adequately powered
Average treatment effect	-10.42 (0.76)	-12.52 (1.20)	-10.42 (0.76)	-7.31 (1.37)	-12.50 (1.20)	-4.39 (0.59)	-10.80 (1.03)	-6.83 (0.61)
SE				-0.97 (0.14)				
SE^2					0.00 (0.00)			

Table B.2: **Average treatment effects, estimated with a variety of models**: The average treatment effects are estimated with random effects models, except for column 6, which is estimated with a fixed effects model. The PET-PEESE models, control for publication bias by controlling for the standard error (SE). Standard errors of the here provided estimates are displayed in parentheses.

	Full sample	Low risk of bias	Adequately powered estimates	Low risk of bias and adequately powered
Chinese pilot ETS	-13.14 (1.04)	-12.93 (1.29)	-8.66 (1.24)	-8.35 (1.33)
EU ETS	-7.27 (1.66)	-7.43 (3.26)	-7.79 (1.38)	-6.17 (2.18)
BC carbon tax	-5.43 (2.15)	-5.32 (2.36)	-5.66 (1.41)	-5.55 (1.42)
RGGI	-21.05 (3.63)	-21.02 (3.74)	-6.10 (5.21)	-6.10 (4.65)
cross-country	-5.85 (3.04)	-4.93 (4.27)	-4.25 (2.10)	-4.65 (2.36)
Tokyo ETS	-7.29 (3.29)	-5.97 (5.96)	-5.46 (2.32)	-5.77 (3.21)
UK carbon price support	-9.78 (3.44)	-7.81 (4.89)	-8.49 (2.43)	-7.02 (3.36)
Saitama ETS	-6.30 (3.51)	-1.06 (5.96)	-5.27 (2.14)	-1.06 (3.21)
California CaT	-18.90 (5.19)		-15.10 (5.21)	
Finnish carbon tax	-14.83 (5.21)	-16.59 (8.58)		
Korea ETS	-3.16 (6.34)		-1.58 (5.17)	
Quebec ETS	-8.54 (4.03)	-10.62 (6.16)	-7.92 (2.95)	-9.38 (3.77)
Swedish carbon tax	-10.17 (4.06)	-5.43 (6.75)		
Australian carbon tax	0.16 (5.83)		0.17 (3.90)	
Swiss ETS	14.25 (12.49)			

Table B.3: **Average treatment effects for individual carbon pricing schemes**: The average effect sizes are estimated with mixed effects models and correspond to the estimates presented in Figure 3.2. The standard errors are displayed in parentheses.

Category	Variables	Description
Carbon pricing scheme	*BC_carbon_tax*, *Chinese_pilot_ETS*, *Finnish_carbon_tax*, *Quebec_ETS*, *RGGI*, *Saitama_ETS*, *Swedish_carbon_tax*, *Swiss_ETS*, *Tokyo_ETS*, *other_schemes*	The dummy variables capture the carbon pricing schemes. EU ETS together with the UK carbon price support are set as benchmark. *other_schemes* includes all other schemes and cross-country studies.
Carbon price level	*log_carbon_price*	Mean carbon price level during studied period. If the primary study reports the carbon prices of the analysed carbon tax or emissions trading scheme, we use these prices. If this information is not provided, we use data by World Bank [200]. For EU ETS we use data by International Carbon Action Partnership [201]. Prices are captured in constant 2010 US$ and transformed to logarithm.
Sector coverage	*industrial_sectors*	The dummy variable captures the sectors energy and industry. Transport and buildings sectors are used as benchmark. For cross-sector studies the variable captures the sectors covered by the policy.
Policy design	*tax*	The dummy variable captures carbon taxing schemes. The benchmark are cap-and-trade schemes.
Study period	*duration*	The variable captures the time between the introduction of the assessed policy and the last included observation.
Study design	*synthetic_control*	The dummy variable captures synthetic control designs, motivated by a previous review, suggesting the relevance of this design [137].
Unit of observation	*Data_Region*, *Data_City*, *Data_Sector*, *Data_Firm*, *Data_Plant*, *Data_Airline*	The dummy variables capture the unit of observation in the studied data. The benchmark is country level data.
Data frequency	*Data_Year*, *Data_Month*	The dummy variables capture the data frequency. The benchmark is data with higher frequency than monthly.
Fuel coverage	*Gasoline*, *Gas*, *Coal*	The dummy variable captures the fuel type, for studies on a specific fuel type. The benchmark are studies across all fuels.
Dependent variable	*DVTotal*	The dummy variable captures studies with total emissions as dependent variable. The benchmark is per capita emissions.
Log transformation	*TransLevelLevel*	The dummy variable captures level-level regression specifications. The benchmark is log-level specification.
Publication bias	*SE_percent*	Capture the standard error of the transformed effect sizes, in line with precision effect test for publication bias [195, 196].
Risk of bias	*Less_Bias*	The dummy variable captures studies with low risk of bias. The benchmark are studies with medium or high risk of bias.

Table B.4: **Moderator variables for the heterogeneity analysis performed in section 3.3.4**: Definitions for the variables used in the heterogeneity analysis. The Bayesian model averaging allows to include a large number of potentially relevant moderators.

	Uniform g-prior				Dilution prior				BRIC g-prior			
	PIP	Post Mean	Post SD	Sign	PIP	Post Mean	Post SD	Sign	PIP	Post Mean	Post SD	Sign
RGGI	0.9995	-28.4470	5.0888	0.0000	1.0000	-30.0563	4.6098	0.0000	1.0000	-30.2213	4.5909	0.0000
Chinese_pilot_ETS	0.9912	-9.7614	2.2285	0.0000	0.9989	-9.6434	1.6511	0.0000	0.9992	-9.6070	1.5812	0.0000
Swiss_ETS	0.8043	14.3507	8.9349	1.0000	0.3463	5.9242	8.8821	1.0000	0.2572	4.3926	8.0679	1.0000
Data_City	0.7759	11.3897	7.6322	1.0000	0.3126	4.2109	6.8207	1.0000	0.2214	2.9425	5.9712	1.0000
duration	0.7614	-0.6351	0.4625	0.0000	0.4002	-0.3015	0.4064	0.0000	0.3519	-0.2731	0.4002	0.0000
synthetic_control	0.4180	2.8697	3.8671	0.9971	0.0385	0.1941	1.1771	0.9282	0.0182	0.0768	0.7402	0.8644
tax	0.4095	-3.1126	4.2424	0.0054	0.1592	-1.4357	3.5543	0.0021	0.1164	-1.0826	3.1861	0.0009
BC_carbon_tax	0.3834	3.8998	5.6509	0.9997	0.1526	1.6532	4.2367	1.0000	0.1113	1.2176	3.7064	1.0000
Swedish_carbon_tax	0.3627	-3.0514	4.6512	0.0171	0.6224	-6.2157	5.2536	0.0002	0.6623	-6.7519	5.2044	0.0001
Coal	0.3250	-2.5837	4.2609	0.0000	0.1232	-1.0401	3.0554	0.0000	0.0865	-0.7370	2.6248	0.0000
Less_Bias	0.3017	1.1583	2.0038	1.0000	0.0559	0.1938	0.9038	1.0000	0.0306	0.1002	0.6485	1.0000
Finnish_carbon_tax	0.2509	-2.8870	5.6951	0.0004	0.1636	-2.0801	5.1188	0.0000	0.1335	-1.7199	4.7358	0.0000
TransLevelLevel	0.1890	-0.6987	1.6730	0.0000	0.0405	-0.1276	0.7251	0.0000	0.0252	-0.0761	0.5564	0.0000
Data_Region	0.1222	-0.3978	1.2963	0.0143	0.0470	-0.1673	0.8779	0.0018	0.0319	-0.1117	0.7179	0.0002
log_carbon_price	0.0880	0.1491	0.6242	0.9883	0.0155	0.0192	0.2230	0.9142	0.0106	0.0127	0.1819	0.9030
Data_Sector	0.0853	0.2481	1.0243	0.9851	0.0206	0.0578	0.5007	0.9925	0.0133	0.0362	0.3954	0.9937
Gas	0.0797	-0.4374	1.8843	0.0000	0.0154	-0.0672	0.7297	0.0000	0.0100	-0.0421	0.5760	0.0000
other_schemes	0.0471	-0.3090	2.0004	0.0000	0.0148	-0.1072	1.1901	0.0000	0.0098	-0.0721	0.9781	0.0000
Tokyo_ETS	0.0431	0.1459	1.0331	0.9966	0.0112	0.0353	0.5134	1.0000	0.0071	0.0219	0.4057	1.0000
industrial_sectors	0.0429	-0.0447	0.7686	0.3686	0.0106	0.0015	0.3050	0.5888	0.0071	0.0038	0.2327	0.6747
Data_Firm	0.0418	0.0709	0.5352	0.9717	0.0122	0.0221	0.2952	0.9876	0.0083	0.0153	0.2445	0.9909
Data_Plant	0.0341	0.0375	0.4701	0.8218	0.0106	0.0163	0.2736	0.9381	0.0074	0.0118	0.2284	0.9595
DVTotal	0.0340	0.0349	0.5058	0.7291	0.0084	0.0043	0.2076	0.6427	0.0062	0.0039	0.1796	0.6788
Saitama_ETS	0.0314	0.0542	0.6530	0.9556	0.0097	0.0193	0.3682	0.9961	0.0064	0.0131	0.2995	1.0000
SE_percent	0.0300	-0.0001	0.0017	0.1833	0.0083	0.0000	0.0009	0.6274	0.0055	0.0000	0.0007	0.7453
Gasoline	0.0300	-0.0158	0.6242	0.4134	0.0083	0.0018	0.2797	0.6259	0.0054	0.0020	0.2210	0.6983
Quebec_ETS	0.0297	-0.0624	0.8224	0.0233	0.0089	-0.0195	0.4548	0.0001	0.0057	-0.0120	0.3608	0.0001
Data_Month	0.0286	-0.0187	0.5163	0.2463	0.0081	-0.0062	0.2653	0.1974	0.0057	-0.0045	0.2209	0.1965
Data_Year	0.0286	0.0105	0.4140	0.6746	0.0081	0.0054	0.2161	0.7795	0.0056	0.0041	0.1794	0.8222
Data_Airline	0.0276	-0.0025	0.7496	0.4954	0.0073	0.0016	0.3707	0.6884	0.0049	0.0017	0.3030	0.8105
(Intercept)	1.0000	-5.9894			1.0000	-6.4805			1.0000	-6.5557		

Table B.5: **Heterogeneity assessment using Bayesian model averaging with alternative priors**: For each of the model specifications and each variable, the table provides the posterior inclusion probability (PIP), the mean (Post Mean) and standard deviation (Post SD) of the posterior distribution for a respective explanatory variable and the share of the meta-regressions where the variable is estimated with a positive sign (Sign). The variables are ordered by their PIP in the first model, which is the main model also presented in the main text.

	PIP	Post Mean	Post SD	Cond.Pos.Sign
log_carbon_price	1.0000	3.8256	0.6633	1.0000
duration	1.0000	-1.5770	0.2594	0.0000
TransLevelLevel	0.9994	-7.1611	1.5894	0.0000
synthetic_control	0.9909	10.2756	2.8058	1.0000
Data_City	0.9618	16.6968	5.8521	1.0000
Coal	0.8677	-10.0338	5.2010	0.0000
Gasoline	0.1481	0.7873	2.3816	0.9999
Data_Airline	0.1382	0.9498	2.8731	1.0000
Data_Plant	0.1233	0.3781	1.2462	1.0000
industrial_sectors	0.1131	-0.3152	1.1226	0.0054
Gas	0.0947	-0.4457	1.7834	0.0000
Data_Sector	0.0868	0.2182	0.9397	1.0000
DVTotal	0.0790	-0.1766	0.8517	0.0022
Data_Month	0.0749	-0.2994	1.6795	0.0326
Less_Bias	0.0661	0.0967	0.5369	0.9989
Data_Firm	0.0621	-0.1098	0.6638	0.0082
Data_Year	0.0572	-0.0917	0.7732	0.0901
tax	0.0473	-0.0060	0.5018	0.6558
Data_Region	0.0450	-0.0007	0.3578	0.6033
SE_percent	0.0432	-0.0000	0.0019	0.0508
(Intercept)	1.0000	-14.3036		

Table B.6: **Heterogeneity assessment using Bayesian model averaging without dummy variables per scheme**: The table provides for each variable the posterior inclusion probability (PIP), the mean (Post Mean) and standard deviation (Post SD) of the posterior distribution for a respective explanatory variable and the share of the meta-regressions where the variable is estimated with a positive sign (Sign). The variables are ordered by their PIP.

Appendix for Chapter 4

Hypothesis
State-owned firms will react less to the carbon price, as their environmental performance is aligned to public emission targets also without the regulation [275, 290] or because the policy is less strictly enforced for state-owned firms [297].
Emission reductions will be lower, if the carbon price regulation is not sufficiently clear and transparent [275].
Carbon pricing policies need appropriate enforcement measures to achieve emissions reductions [298, 389].
For ETSs the possibility to bank emission allowances across years can foster emissions reductions in periods of relatively high supply of allowances [240, 254].
Emissions reductions are influenced by how emission allowances are allocated to the firms and to what extent this is exogenous to the firms decisions [174, 269, 271, 272, 275, 279, 297].
Higher carbon prices lead to higher emissions reductions [18, 178, 263, 271, 277, 279].
Large-scale firms reduce emissions more, as installations for emissions reductions are more profitable for larger facilities due to scale effects [291].

Table C.1: **Hypotheses not assessed in the realist synthesis**: The table lists hypotheses, expressed in primary studies, which could not be captured in the assessment of the realist synthesis, as they could not be allocated to a specific mechanism.

Code	Assigned elements
Context	
Fuel price	5
Electricity price	1
Business cycle	8
Emission trend	33
Other Policies	15
Alternatives available	2
Abatement costs	9
Firm characteristics	9
Financial constraints	7
Ownership	6
Firm size	5
Sector	18
Buildings	15
Transport	32
Aviation	12
Industry	43
Energy	65
Market structure	7
Merit order	3
Policy design	22
Coverage	2
Off-sets	1
Enforcement	8
Overallocation	4
EU ETS phase I/II/III	18
Banking/borrowing	2
Allocation mechanism	38
Price level	51
Mechanism	
Electrification	4
Low-carbon technology	14
Investment	11
Anticipation of future prices	4
Salience	11
Innovation	34
Reduction coal use	26
Announcement effect	17
Fuel switch	71
Efficiency	65
Energy intensity output	6
Carbon intensity energy	6
Carbon intensity output	17
Other	9
Leakage	58
Reduced production/consumption	56
Switch to alternatives	4
Type	
Evidence	253
Discriptive	34
Qualitative	6
Hypothesis	177

Table C.2: **Coding scheme for the data extraction for the realist synthesis**: The codes were created iteratively during the coding process. They are here presented together with the number of elements that were assigned this code during the data collection. The categories were further synthesised during the analysis of the data, as they are reported in Chapter 4. Indented codes are subcategories of the higher-level code, but their occurrence is only counted in the subcategory to avoid double counting.