



GREEN
CLIMATE
FUND

Independent
Evaluation
Unit



EVIDENCE REVIEW ON BEHAVIOURAL CHANGE IN DEVELOPING COUNTRIES

Approach paper

Samantha Booth, Romina Cavatassi, Benjamin Curtis, Deborah Sun
Kim, Laurenz Langer, Caitlin Blaser Mapitsa, Promise Nduku,
Martin Prowse, Jyotsna Puri, Jamie Robertsen, Carina van Rooyen

February 2022

Evidence review on behavioural change in developing countries

APPROACH PAPER

Samantha Booth, Romina Cavatassi, Benjamin Curtis, Deborah Sun Kim,
Laurenz Langer, Caitlin Blaser Mapitsa, Promise Nduku, Martin Prowse,
Jyotsna Puri, Jamie Robertsen, Carina van Rooyen

02/2022

CONTENTS

LIST OF AUTHORS	IV
ACKNOWLEDGEMENTS	IV
A. BACKGROUND	1
1. Description of the problem.....	1
2. Why is it important to do this review?	2
B. OBJECTIVES OF THE REVIEW	4
C. METHODS	4
1. The overall systematic review design approach.....	4
2. Theory of change.....	5
3. Intervention/outcome framework for the EGM	7
4. Criteria for inclusion and exclusion of studies in the review	13
5. Searching for evidence.....	16
6. Data collection and analysis.....	22
D. DISTRIBUTION OF MAIN ROLES.....	28
REFERENCES	30
APPENDICES.....	33
Appendix 1. Theory of Change (ToC)	34
Appendix 2. Intervention-outcome framework	35
Appendix 3. Summary inclusion and exclusion criteria organized using the PICOS (population, intervention, outcome, comparison, study design) model	42
Appendix 4. Search terms for each sector	47
Appendix 5. Data extraction tool.....	50
Appendix 6. Critical appraisal tool.....	56
Appendix 7. Grading of recommendations assessment, development and evaluation (grade) tool (example)	61

TABLES

Table 1.	Behavioural intervention definitions	9
Table 2.	Outcome definitions	11
Table 3.	List of targeted databases	17

FIGURES

Figure 1.	Theory of change	6
-----------	------------------------	---

LIST OF AUTHORS

The authors of the review of the evidence review on behavioural change in developing countries are (in alphabetical order by last name):

FULL NAME	AFFILIATION
Samantha Booth	Human Friendly Solutions
Romina Cavatassi	International Fund for Agricultural Development
Benjamin Curtis	Human Friendly Solutions
Deborah Sun Kim	Independent Evaluation Unit, Green Climate Fund
Laurenz Langer	Africa Centre for Evidence, University of Johannesburg
Caitlin Blaser Mapitsa	School of Governance, University of the Witwatersrand
Promise Nduku	Africa Centre for Evidence, University of Johannesburg
Martin Prowse	Independent Evaluation Unit, Green Climate Fund
Jyotsna Puri	International Fund for Agricultural Development
Jamie Robertsen	Independent Monitoring, Evaluation and Learning Consultant
Carina van Rooyen	Africa Centre for Evidence, University of Johannesburg

ACKNOWLEDGEMENTS

Any errors or inconsistencies are entirely the responsibility of the authors.

A. BACKGROUND

1. DESCRIPTION OF THE PROBLEM

Climate change is projected to intensify over the next several decades, resulting in a host of negative impacts on socio-ecological systems. The need for both adaptation and mitigation is clear. Equally clear is that human behaviour is a key driver of climate change and that many adaptation and mitigation strategies require changes to behaviour. Where rigorous empirical guidance is lacking, however, is how to change behaviour most effectively to support adaptation and reductions in emissions. In particular, research evidence relating to developing countries is somewhat thin and scattered. This is a pressing problem given that the negative impacts of climate change will not be uniformly distributed across the globe: developing countries are likely to be disproportionately affected. Therefore, understanding what is effective in changing behaviour in these countries is particularly important.

The Global Commission on Adaptation flagship report emphasizes the need for immediate action to anticipate economic, environmental and humanitarian costs of potential disruption (Global Commission on Adaptation, 2019). Climate change temperature and climate have already changed considerably. This shift is projected to intensify over the next couple of decades and there is no doubt these changes are largely driven by human behaviour (Solomon and others, 2007).

The impacts of climate change will not be uniformly distributed across the globe. If a broad brush is applied to climate change, developing countries are more likely to disproportionately experience the negative effects of global warming. Due to geographical locations, developing countries tend to have warmer climates than those in the developed world, and they are also reliant on climate-sensitive sectors such as agriculture, forestry and tourism. As temperatures continue to rise, regions such as Africa will face a decline in crop yields and will struggle to produce adequate food for consumption, whilst their key exports will likely decrease in volume (Wade and Jennings, 2015). Adaptation planning efforts by the United Nations Framework Convention on Climate Change (UNFCCC) through national adaptation programmes of action (NAPAs), nationally determined contributions (NDCs) and national adaptation plans (NAPs) have played a key role in ensuring resources are available for countries to articulate climate adaptation needs and resume implementation (Binet and others, 2021).

The primary driver of current climate change is greenhouse gas (GHG) emissions from human behaviours, such as the burning of fossil fuels. Hence, climate change may be mitigated by changes to these behaviours. Yet, human behaviour is the least-understood aspect of the climate change system (Bernstein and others, 2008), albeit the main driver of the problem. In this respect, climate change is a key concern of cognitive processes, particularly covered by environmental psychology, behavioural economics, and behavioural science (Schmuck & Schultz, 2012). The literature highlights a complex set of interrelated psychological factors that hamper climate change and action (Gifford and others, 2011; Stoknes, 2014; Van der Linden and others, 2015), such as perceived distance, framing and cognitive dissonance (Stoknes, 2014).

In recent decades, theories and evidence from behavioural science – defined by Balmford and others (2021) as the scientific study of behaviour, informed by an array of disciplines including sociology, psychology, economics, anthropology, and political science – have provided insight into the social, motivational, cognitive, cultural and contextual factors underlying human behaviour. According to Stern (2020), behavioural interventions involve neither command and control regulations nor financial incentives. Examples include information provisions, appeals to values and norms,

engagement, and restructuring choice options (commonly referred to as nudges). These insights have informed interventions that have helped to encourage societally valued behaviour change including reductions in smoking, addiction and obesity, as well as improvements in tax compliance, development assistance and climate change mitigation (Duflo and others, 2011; Datta & Mullainathan, 2014, Hallsworth and others, 2017; Bollinger and others, 2020). Research has informed behaviour change interventions relevant to a variety of environmental issues including, but not limited to, energy efficiency, water conservation, recycling and transport (Osbaldiston & Schott, 2012; Byerly and others, 2018; Nisa and others, 2019).

We have an opportunity and a responsibility to affect change through increased understanding of the factors that underlie anthropogenic causes of climate change and the ways in which mitigation and adaptation behaviours may be encouraged effectively (Gifford and others, 2011). A body of knowledge in behavioural science illuminates some key mechanisms that underlie climate-relevant behaviour, and indicates some promising avenues for human responses to climate change. Aiming to fill the knowledge gap in climate action, behavioural science identifies promising ways with which to increase the effectiveness and impact of climate investments (Krüger & Puri, 2020). Insights from behavioural science have been frequently applied to enhance public policy effectiveness (OECD, 2017). For example, nudges, a category of psychology-based interventions, can be a cost-effective tool for supporting individual decision-making and have been applied to foster pro-environmental behaviours (Cinner, 2018; Schubert, 2017). Nudges can involve simple alterations to the physical micro-environments in which choices are made (choice architecture). Such small changes can have significant effects on behaviour, helping people to make decisions more beneficial for themselves and the broader society (Szasz and others, 2018; Thaler & Sunstein, 2009; Hollands and others, 2017).

Bamford and others (2021) argue that integrating evidence from behavioural science into the design of biodiversity conservation interventions currently based on education, regulation, and material incentivization, shows great potential for enhancing their effectiveness (Bamford and others, 2021). Traditional interventions in conservation campaigns try to persuade consumers, farmers or politicians to change their behaviour by highlighting the environmental impacts of their actions, but these broad-breadth attempts to increase knowledge are often not sufficient to shift behaviour (Kollmuss & Agyeman, 2002). Effective communication campaigns for global issues like climate change or pandemics are proven to be two-way processes that involve clear messages tailored for diverse audiences. They are shared by trusted people and incorporate actions by individuals that demonstrate a clear contribution to addressing the problem (Hyland-Wood and others, 2021). Behavioural science also shows that information campaigns can be more effective when they target discrete audience segments and account for their values as well as social and physical realities (Cheng and others, 2011; Kahan and others, 2012; Kusmanoff and others, 2020). For instance, switching from pro-social to self-interest messages is seen to increase the adoption of solar panels in the United States (Bollinger and others, 2020). Changing default settings for sign-up documentation in such a way that participants must opt-out (rather than into) individually or societally more desirable choices proved to be an effective strategy that increased household subscriptions to renewable energy programmes (Eleking & Lotz, 2015; Liebe and others, 2021).

2. WHY IS IT IMPORTANT TO DO THIS REVIEW?

This review aims to examine which interventions are effective in promoting environmental and development outcomes from individuals, households, communities and companies, and how effective efforts have been to date. The specific focus of the review is driven by the growing hope that behavioural interventions (Schot and others, 2016; Stern and others, 2016) such as consumption feedback, social comparison messages or tailored environmental appeals may potentially be cost-

effective strategies that can complement traditional market tools and regulation. To the best of our knowledge, there appears to be an absence of systematic analyses that carefully explore the nature of evidence and the relative effectiveness of behavioural science interventions on environmental and development outcomes, particularly in developing countries. In summary, there is extensive evidence about both what is ineffective, and what works in promoting behaviour change broadly (Flanagan & Tanner, 2016), but it has not been rigorously mapped or synthesized in the climate sector in developing countries specifically. This review will reduce the gap within the literature, to inform governments, donors and other decision-makers of the available evidence on a broad set of behavioural science interventions and their impacts on climate change mitigation and adaptation (including human development) across different sectors in developing country contexts.

Through this review, we will be able to rigorously map what evidence currently exists for climate-related behaviour in developing countries, and precisely synthesize selected bodies of evidence to discern what behavioural science interventions work for attaining desired environmental and development outcomes. Drawing on behavioural science knowledge in developing country contexts, we will conduct a meta-analysis of behavioural science interventions that are most effective in less developed countries for changing climate-related behaviours that include both actions for mitigation and for adaptation. We contribute to the literature of behavioural science interventions in the following ways:

- We synthesize several models of behaviour change to create a new categorization of behavioural interventions (see section C3), which is the foundation of this review and is also validated by an overall theory of change (ToC). Whilst the review is broad in scope, we have a precise but extensive list of interventions and outcomes with clearly structured categories. This allows us to search for evidence on behavioural science interventions across fields of studies but at the same time, keep the scope of the review manageable.
- To learn about the causal evidence for behavioural science interventions on environmental and development outcomes, we select only quantitative studies with experimental and quasi-experimental study designs. The inclusion criteria is based on a precise definition of behavioural science interventions (see Table 1) and include studies that have the potential to document environmental and development outcomes according to this criteria. We include studies where data was collected at any reasonable point after the intervention.
- At the first stage, we provide a framework of reviewed evidence in the form of an evidence gap map (EGM) of behavioural science interventions across different sectors. An EGM is a convenient and user-friendly tool for policymakers to quickly inform themselves about the existing evidence. Through this exercise, we will be able to highlight areas where research is comprehensive and where evidence gaps exist. Additionally, this will enable policy makers and practitioners to make informed decisions about project prioritization and commissioning of further research activities.
- We then conduct a meta-analysis with data extracted from selected quantitative studies of sufficiently populated cells of the EGM (i.e. at least 10 studies from the same intervention and outcome combination). This exercise is not common in the literature on changing climate-related behaviours.
- The results of the meta-analysis are important for determining where robust evidence exists, across individual studies and contexts, for behavioural science interventions on environmental and development outcomes, and for establishing what the synthesis of aggregate effects tells us about the magnitude and direction of impact. This will minimize the risk that any large effects of interventions are simply outliers and will also help us assess the confidence with which these effects are measured.

B. OBJECTIVES OF THE REVIEW

Our review seeks to answer the following review question:

- To what extent are behavioural science interventions conducted in developing countries effective in producing environmental and development outcomes? In answering, we address the following review objectives:
- To produce an interactive EGM on research evaluating the effectiveness of behavioural science interventions on climate, environmental and development outcomes in developing countries.
- To provide a rigorous synthesis of causal evidence to identify the effects of behavioural science interventions in supporting behavioural change on climate, environmental and development outcomes in developing countries.

C. METHODS

1. THE OVERALL SYSTEMATIC REVIEW DESIGN APPROACH

We aim to use a two-stage systematic review (SR) approach that consists of a first-stage EGM and a second-stage SR and synthesis, in compliance with guidelines for the production of EGMs and SRs outlined by the Campbell Collaboration.¹ For instances in which these guidelines do not cover the linking of the two evidence products, we propose an effective and adaptable research method that fully integrates the EGM and SR processes. Suggestions based on previous synthesis project experiences in the environmental sector – such as ecosystem services for poverty alleviation (Erasmus and others, 2017), payment for ecosystem services (Snilstveit and others, 2018) and gender (Langer and others, 2018) – indicate that the successful integration of an EGM and subsequent full SR is dependent on four key factors:

- Continued and embedded stakeholder engagement on the scope of the overall project and both synthesis outputs
- A consistently rigorous and transparent synthesis approach that applies similar criteria of rigour to both outputs
- A sufficiently broad scope and design of the EGM that guarantees a sufficient evidence-base for subsequent synthesis
- A versatile software solution to provide flexibility in the backend as well as the frontend of the evidence mapping tool applied to integrate the knowledge management aspect of the evidence review, with the visualization requirements of the EGM

These factors inform our overall methodological approach to this project and have implications for the structure of the research process with two elements requiring upfront design before discussing the detailed methodologies for the EGM and the SR respectively.

a. Evidence gap map

The EGM will have a broader scope in terms of inclusion of evidence than the full SR, but both are focused on the nature of the evidence that exists regarding the effectiveness of behavioural science interventions on climate, environmental and development outcomes in developing countries. The EGM will map evidence from impact evaluations and SRs on behavioural science interventions. Its

¹ For SRs, see Campbell Collaboration (2020). [Campbell systematic reviews: Policies and guidelines](#), accessed on 16 December.

For evidence and gap maps, see Campbell Collaboration (2020). [Guidance for producing a Campbell evidence and gap map](#), accessed on 19 November.

main objective will be to indicate the overall nature and size of the available evidence base, and to identify areas for synthesis and substantiate evidence gaps for future commissioning.

The EGM will apply an intervention-outcome matrix to structure the identified evidence-base and, by doing so, will highlight the size and nature of the evidence for the different configurations of interventions and environmental and development outcomes. The EGM will be visualized on an interactive online interface using Evidence for Policy and Practice Information and Co-ordination Centre (EPPI) mapping software,² similar to the International Initiative for Impact Evaluation (3ie) Evidence Maps.³ This mapping software allows for multiple visualization options as well as integrated user feedback in addition to a range of other minor advantages above other software solutions. In consultation with the engagement committee, we have the option to produce a series of EGMs using different segmenting attributes to foreground different attributes of the evidence base. That is, the visualization and colour patterns can be altered to divert from the traditional EGM approach which foregrounds the impact evaluation/SR distinction as well as the quality of reviews. Alternative attributes to foreground could refer to the gender sensitivity of the included evidence base or implementation considerations.

Stakeholders will also be able to use the interface, depending on preference, to create customized maps by filtering the evidence base according to attributes such as region and study design. The EGM is a product in its own right that supports stakeholder engagement with the evidence base and also supports decision-making on the most effective synthesis approach and scope. The evidence base included in the map will aid the identification of the interventions and outcomes that are of most interest to stakeholders and policy-makers. It will be applied instrumentally to guide discussions on which areas of the evidence base to use for synthesis, as well as on the most effective method for synthesizing the evidence in answering the review question.

b. Systematic review and synthesis

Following the completion of the EGM, we will conduct an SR and synthesis of selected bodies of evidence contained in the EGM. This can refer to selected cells within the EGM or selected intervention categories or outcome variables depending on stakeholder preferences. An effectiveness review will be conducted to answer the review question on the extent to which interventions have been effective at achieving behavioural change in desired environmental and development outcome areas in developing countries. Therefore, the SR will only include primary studies that measure the effects of interventions and have designs that can reliably attribute observed effects to the applied interventions. Individual effects will be synthesized into overall estimates of treatment effects using statistical meta-analysis.

2. THEORY OF CHANGE

In the context of the evidence review on behavioural change, the purpose of the ToC is to inform the types of interventions included in the EGM. The ToC directly informs the Population, Intervention, Comparator, Outcome and Study (PICOS) design framework that will be used to develop inclusion and exclusion criteria. A ToC is, essentially, mechanisms (i.e. the how and why) "through which results it aims to affect (Frey, 2019, p. 1315). In the context of the evidence review on behavioural

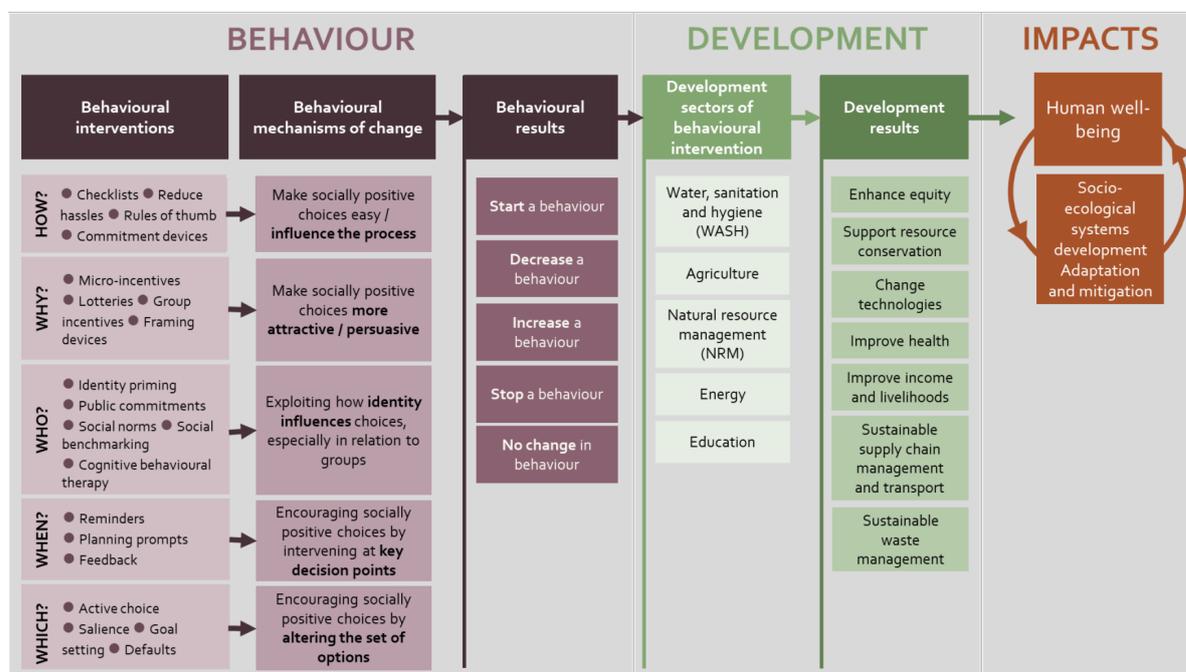
² For an example, see Africa Centre for Evidence (2020). *Engaging stakeholders with evidence & uncertainty: an evidence map*, accessed in June.

³ For an example, see International Initiative for Impact Evaluation (2017), *Social, Behavioural and Community Engagement Interventions for Reproductive, Maternal, Newborn, Child Health*, accessed in March.

change, this ToC illustrates the role that behavioural science interventions play in human and environmental development outcomes, and is shown in Figure 1 (see also Appendix 1).⁴

The ToC is divided into three distinct parts – behaviour, development, and outlines three levels of intervention, while “Impact” is the culmination of the process. Within behaviour, the ToC conceptualizes three levels showing how behavioural interventions lead to behavioural results, through five behavioural mechanisms of change. Definitions of the behavioural interventions are provided in Table 1 below (also see intervention-outcome section C3). The first level is a categorization of different behavioural interventions (such as checklists, social norms, defaults, etc.). These interventions are the ones most commonly applied in the field, and are drawn from the list compiled by the Behavioural Science Evidence Hub, a leading knowledge clearinghouse for policy-relevant behavioural science. The second level specifies the mechanisms of change, that is, how these interventions actually influence behaviour (such as through changing sets of options or ‘nudging’ at key decision points).

Figure 1. Theory of change



Source: Authors (see Appendix 1 for a larger image)

These mechanisms are informed by two prominent conceptualizations of behaviour change: the Easy Attractive Social Timely (EAST) framework produced by the Behavioural Insights Team (Service and others, 2014), and the 4Ps [possibilities, process, persuasion and person] framework created by Yale University (Dhar, 2014). The third level outlines concrete behavioural results (e.g. starting a behaviour, stopping a behaviour, etc.).

⁴ The ToC also is expected to evolve based on the outcomes of this review. As such, it is suggested that the categories presented are suggested may be further developed to reflect the analytical requirements of the study. In using the ToC to inform the evidence review, it is possible that certain additional categories will emerge, based on the results of the search. This could include the addition or division of the sectoral component of the ToC. Furthermore, traditional development programming around behaviour often includes components of knowledge and attitude. These have been included in the matrix, to support the search process and better understand the relationships between newer behavioural science strategies, and traditional development models. They have not been included in the ToC, because they do not reflect our current understanding of mechanisms of change.

This ToC is unique in that the outcome of the behavioural intervention leads to activities that are an input for the development component of the ToC. There are therefore two intervention levels before results are attained in human and environmental development. These development results, which are grouped by indicative sectors, then have their own intermediate and wider outcomes. The development results are purposefully categorized more broadly than the behavioural change interventions and results. This is to ensure that, (i) the ToC is not so complex as to lose utility, and (ii) that the ToC does not limit the development results in the evidence gap mapping process. The move, in the ToC, from narrowly defined interventions to broad development results, also means that the causal pathways are less well-articulated. Two examples that outline potential causal pathways are provided below:

- If the development result was to adopt new farming practices (improve income and livelihoods) through an agriculture intervention, the potential causal pathway would be using planning prompts (behavioural interventions) to encourage socially positive choices by intervening at key decision points (behavioural mechanism of change). This would then result in starting a behaviour (behavioural result), which in this example would be adopting new farming practices.
- If the development result was to use more energy-efficient lighting (change technologies), through an energy-related intervention, the potential causal pathway would be using micro-incentives (behavioural interventions) to make positive choices more attractive/persuasive (behavioural mechanism of change). This would then result in starting a behaviour (behavioural result), which in this example would be using energy-efficient lighting.

The development sectors were selected based on their potential for behavioural interventions with outcomes that have results impacting socio-ecological systems. It is possible that, after the completion of the EGM, it will emerge that some sectors either need to be divided for more granularity or that they do not actually contain relevant studies. Since the ToC is a living document, it can reflect the ongoing findings of the review. Similarly, the development results have been selected as part of an iterative process of refinement. They are not necessarily mutually exclusive, and one intervention could target more than one result area. These are described in Table 2 of the intervention-outcome section C3 below.

The impact level denotes the desired state of stable socio-ecological systems through human wellbeing, and climate change adaptation and mitigation. Crucially, these two impacts are intrinsically linked, but for the purposes of this study, we will consider the intention of the research when determining contribution towards impact. Following on from the causal pathway examples provided above:

- Adjusting farming practices to new climate conditions contributes to climate change adaptation and improves human wellbeing through sustaining or improving incomes and livelihoods. This in turn contributes to developing and sustaining more stable socio-ecological systems.
- Changing technologies by using energy-efficient lighting contributes directly to mitigating the effects of climate change by reducing energy consumption. This in turn contributes to developing and sustaining more stable socio-ecological systems.

3. INTERVENTION/OUTCOME FRAMEWORK FOR THE EGM

The EGM intervention-outcome framework is the primary tool for structuring and visualizing the evidence base, and its design is directly influenced by the ToC above. Appendix 2 illustrates the structure of the intervention-outcome framework for the EGM. The dimensions of the map are placed in a matrix format of row and column headings that are used to structure the evidence base. The primary dimensions of the EGM are intervention categories (row attributes) and the outcome

domains (column attributes), which are divided into subcategories and subdomains respectively. The structure of our intervention-outcome framework maps the key behavioural science interventions onto outcomes broadly divided into attitudes (intermediate outcome), behaviour (final outcomes), development (development results), and impact (socio-ecological systems development including human wellbeing). Definitions of the behavioural interventions and outcomes are provided in Table 1 and Table 2, respectively.

Table 1. Behavioural intervention definitions

BEHAVIOURAL INTERVENTION	DEFINITION
HOW is the choice made? This category of intervention influences the decision-making process to make positive choices easier.	Checklists This type of intervention creates a series of procedural steps to guide decisions or behaviour. The steps are designed to be used consciously and systematically, and thereby reduce the complexity of decisions/behaviour.
	Reduce hassles This type of intervention removes procedural or processual barriers standing in the way of positive behaviours. Reducing hassle and barriers means there is less "friction"
	Rules of thumb This type of intervention simplifies decision-making by creating a relatively straightforward heuristic device. In distinction to checklists' cognitive, relating to how people think about rules and decisions rather than how they carry them out.
	Commitment devices In this type of intervention, people consciously commit to following a certain course of action/behaviour. The specific "device" takes salvability of forms, but it typically encouraging positive decisions in the present.
WHY is the choice made? This category makes positive choices more attractive or persuasive.	Micro-incentives This type of intervention typically involves small rewards given out to encourage specific behaviours. The incentives are often but not exclusively cash, can be frequent, and are tied to the completion of tasks.
	Group incentives This type of intervention gives rewards based on a group group members all complete a designated behaviour, then the entire group receives the reward.
	Lotteries A lottery encourages a positive decision by holding out the promise of some reward in the future. Even if the probability of winning the reward is small, it can incentivize behaviour.
	Framing devices A framing device influences decisions via often subtle changes in how the options are presented. Certain options are made to seem either more or less attractive through highlighting potential loss, gain, or risk, which are three common, potential "frames".
WHO is making the choice? This category of interventions exploits how identity influences decision-making, especially in relation to groups, to encourage environmentally positive choices	Identity priming This type of intervention influences decision-making, particularly in relation to group memberships. "Priming" involves exposing an individual to a concept or cue that influences subsequent behaviour. In practice, personal, civic decisions to encourage the individual to take actions consistent with ostensible group values.
	Public commitments This type of intervention is a commitment device in which people promise to others that they will take a certain course of action/behaviour. Other individuals or the group thereby hold the individual accountable for his/her behaviour.
	Social norms This type of intervention leverages an individual providing information on what is commonly accepted or disapproved (such as 'approved' or 'disapproved' norms) to encourage/

BEHAVIOURAL INTERVENTION		DEFINITION
	Social benchmarking	This type of intervention directly compares an individual's behaviour using measurable data (such as energy consumption) to a target behaviour.
	Cognitive behavioural therapy	This is a therapeutic intervention that influences behaviour by getting people to think about their thinking. It typically provides a structure to alter thought patterns that give rise to certain behaviours.
WHEN is the choice made? This category of intervention encourages positive choices by influencing key decisions.	Reminders	This type of intervention involves messaging people (via email, SMS, etc.) in a timely way to call their attention to something, and/or to encourage them to take certain actions.
	Planning prompts	In this type of intervention, people are prompted to plan for when, where and how they will undertake certain actions. The prompt typically helps them think through a process for deciding and then carrying it out, and can frame future benefits of the behaviour in a more short-term timeframe.
	Feedback	This type of intervention provides information, often tracked over time, about behaviours. The information might report how the tracked behaviours compare to targets, and/or outline consequences of the behaviour trajectories.
WHICH choices are available? This category encourages positive choices by altering the set of options.	Active choice	This type of intervention makes clear which of a series of options will lead to a better outcome. It forces a choice because there is no default and highlights potential losses from choosing the less-desirable option(s).
	Saliency (communication)	This type of intervention improves the ease and accessibility of adopting behaviours by making information/choices more prominent and relevant. Personalizing communication and highlighting follow-on instructions are typical strategies to increase saliency. Because it focuses on messaging content rather than timely delivery, it is distinct from a reminder.
	Saliency (experience design)	This type of intervention targets how individuals interact with their physical and/or digital environment. It involves arranging facilities or options so that they are either: (i) more prominent, accessible, and easy, to prompt a particular behaviour or, (ii) Less prominent, accessible, or easy, to discourage a particular behaviour.
	Goal setting	This type of intervention helps individuals consider what their priorities are, then specify a series of goals that they would like to achieve. It often goes along with a planning process.
	Defaults	This type of intervention involves setting a default option that people must actively choose to change. The default is typically set as the socially optimal choice, encouraging people to stick with that option.

Table 2. Outcome definitions

OUTCOME		DEFINITION
Intermediate outcomes (knowledge and attitudes)	Knowledge of intervention	Awareness of the intervention and its objectives.
	Take part in intervention	Adoption of intervention activities.
	Acquire knowledge	Increased understanding of environmental and development related issues.
	Change attitudes	Perceptions on the environment and developmental matters.
Final outcomes (behaviour change)	Start behaviour	Resumption of actions/activities following the intervention.
	Increase behaviour	Evidence of more actions/activities due to the intervention.
	Decrease behaviour	Reducing actions/activities.
	End behaviour	To halt actions/activities.
	No change in behaviour	No evidence of noticeable variations from the status quo regarding conduct.
Development results	Enhanced equity	Inequality often drives unsustainable systems of production and consumption, and many interventions aim to share resources in a community more equitably.
	Natural resource conservation and preservation	This result could include outcomes such as reduced water use, a reduction in the harvesting of wild plants, limiting encroachment on protected areas, or the improvement of soil quality.
	Changed technologies	This result includes an evolution in technology used, such as more drought-resistant seeds, improved cooking stoves, or water-efficient toilets.
	Improved health	While health is not a core sector being included in the review, many interventions in sectors such as water, sanitation and hygiene (WASH), agriculture, and transport, have aims of improving health; this is a key component of wellbeing. Results could include improved nutrition or a reduction in illnesses linked to air pollution or water quality.
	Improved income or livelihoods	The interlinkages between income and ecological outcomes is complex, but many human development interventions have an increase in income as a key result.
	Sustainable transport or supply chain management	This result will include transport options that reduce fossil fuel consumption, reduce private vehicle ownership, or increase uptake in public transport, strengthen transport management systems, or support local suppliers.
	Sustainable waste management	This result will reflect interventions promoting separation at source, reduced packaging, composting, and other waste-related practices.

OUTCOME		DEFINITION
Socio-ecological systems development (includes human wellbeing)	Mitigation	Shift to low-emission sustainable development pathways (human wellbeing). Examples: Increased low-emission energy access and power generation Use of low-emission transport Reforestation, sustainable forest management, afforestation, agroforestry practices Low- or zero-carbon livestock Zero or minimum tillage, sustainable rice intensification Reduced emissions from buildings, cities, industries and appliances Strengthened institutional and regulatory systems for low-emission planning and development
	Adaptation	Increased climate-resilient sustainable development (human wellbeing). Examples: Increased resilience of infrastructure and the built environment to climate change threats Increased generation and use of climate information in decision-making Strengthened adaptive capacity and reduced exposure to climate risks Strengthened institutional and regulatory systems for climate-responsive planning and development Adoption of adaptation options promoted by the intervention (use of climate-resistant varieties, conservation agriculture, sustainable rice intensification, rotational plans for pasture and fishery, etc.)

4. CRITERIA FOR INCLUSION AND EXCLUSION OF STUDIES IN THE REVIEW

To systematically characterize a large, disparate literature on the effectiveness of behavioural science interventions, an underlying focus on environment and human development outcomes guides the scope of the review. Formally, we adopt the PICOS design framework to develop our inclusion criteria. Summaries of the inclusion criteria for the EGM and full SR are provided in I. Annex 1. Appendix 3.A and Appendix 3B, respectively. The inclusion criteria define the precise characteristics of the studies included in the review. All evidence not meeting these criteria will be excluded as it is beyond the scope of the review.

The inclusion criteria for the EGM and the SR overlap completely, apart from two exceptions:

- The EGM includes impact evaluations and SRs while the full SR only includes impact evaluations. All SRs included in the EGM will be extracted and considered for inclusion in our own review.
- The scope of the SR will include a selected number of the interventions and outcomes configurations provided in the EGM. That is, post the completion of the EGM, we will engage with all stakeholders to settle on the most relevant intervention and outcomes configurations for synthesis. This decision depends on the extent of the available evidence base and decision-makers' evidence needs.

a. Population

We follow the country-level categorization of developing countries in the UNFCCC Kyoto Protocol, and include studies assessing the effectiveness of a behavioural science intervention in:

- Non-Annex 1 countries⁵
- Non-Annex 1, and Annex 1 countries jointly if the associated analysis distinguishes effects and reports results separately across the two samples

Any primary study that presents combined analysis on both Annex 1 and non-Annex 1 countries without reporting separate results across the two samples will be excluded. Systematic reviews are included in the EGM either if data is aggregated for non-Annex 1 countries relative to Annex 1, or if there is at least a single primary study included that is from non-Annex 1 countries.

The EGM and SR will include studies conducted at any unit of observation, for example including individuals, households, communities and companies. We will consider studies published only from the year 2000 onwards.

b. Interventions

We will only include behavioural science interventions, which are all informed by empirical research principally from behavioural psychology and/or behavioural economics. Relevant research seeks to identify characteristic human cognitive patterns which are often unconscious or not "rationally maximally economic sense." Building on these patterns, interventions can alter the choice architecture of decision-making, building "nudges" to overcome barriers, and optimize communications, all with the typical goal of encouraging pro-social behaviours. The type of interventions we include are informed by the ToC described in section I. Annex 1. Appendix 1.C.22. Table 1 above presents an overview and definitions of the behavioural interventions that will be included.

⁵ UNFCCC (2020). *Parties to the Convention and Observer States*, accessed in August.

Interventions can be delivered at any administrative level, and administered to any type of beneficiary (e.g. individual, household) by any type of actor (e.g. government, non-governmental organization). Additionally, we will not put any restrictions related to intervention-level characteristics such as modality, intensity, duration, or complexity of intervention delivery. Specifically, we will not exclude studies based on restrictions related to sample size, ensuring that pilot-scale interventions that often focus on newer, more innovative approaches are captured in our evidence review.

c. Comparison

The review will only consider evaluation studies that clearly identify at least two experimental groups: (i) a treatment group exposed to the intervention and (ii) a control group that does not receive the intervention for the purpose of establishing the impact of the intervention. The nature of the control group will depend largely on the specific methods deployed in the study (e.g. the control group in a randomized controlled trial) and can refer to the population receiving no treatment, treatment as usual, placebo treatment, or pipeline treatment. We will consider synthetic control groups for inclusion.

We will exclude any study that does not describe a clearly articulated control group, for instance, descriptive/predictive analyses highlighting drivers and determinants of selecting into behavioural science interventions. Quantitative methods for which the use of comparison/control groups is not relevant, such as life-cycle assessments, will be excluded.

d. Outcomes

Studies will have to assess the effect of the above interventions on the following final and/or intermediate outcomes, development results, and impacts outlined in section C3 above to be included in the EGM and SR. The EGM will consider the following intermediate outcomes, final outcomes, development results and impacts as shown in the ToC (see Figure 1).

We will assess these ranges of outcomes measured at any unit of analysis (e.g. individual, household, community, and organizational level). Moreover, in line with our broad criteria related to study-level characteristics, we will consider studies that measure outcomes at any reasonable point following the administration of the relevant behavioural science intervention. We are not pre-specifying relevant outcome indicators for inclusion of studies in this review, and we will assess the validity of indicators and outcome measures as part of the critical appraisal of evidence. We will also record information on unintended outcomes, for example an increase in environmental degradation, as well as information on intervention costs or cost-effectiveness where reported.

e. Study design

We will include SRs and impact evaluations in the EGM with the following definitions and designs specifying both study types. For our SR, only impact evaluations will be included, using the same definition and designs as for the EGM.

1) Systematic reviews eligible for inclusion

We will include any form of literature review or evidence synthesis – regardless of whether or not it self-identifies as an SR – as long as the review describes its search for evidence, data collection, and methods for synthesis.⁶

2) Impact evaluation designs eligible for inclusion

⁶ This follows Snilstveit and others (2016) and overlaps with 3ie's inclusion criteria on the Evidence Review Portal.

We will include studies that assess the effects of interventions using experimental designs or quasi-experimental designs with non-random assignment that allow for causal inference, in line with Lwamba and others (2020). Specifically, we include the following:

- a) Randomized controlled trials (RCTs), with assignment at individual, household, community, or other cluster level, and quasi-RCTs using prospective methods of assignment such as alternation.
- b) Non-randomized studies with selection on unobservables:
 - i) Regression discontinuity designs, where assignment is done on a threshold measured at pre-test, and the study uses prospective or retrospective approaches of analysis to control for unobservable confounding.
 - ii) Studies using design or methods to control for unobservable confounding, such as natural experiments with clearly defined intervention and comparison groups, which exploit natural randomness in implementation assignment by decision-makers (e.g. public lottery) or random errors in implementation, and instrumental variables estimation.
- c) Non-randomized studies with pre-intervention and post-intervention outcomes data in intervention and comparisons groups, where data are individual-level panel or pseudo-panels (repeated cross-sections), which use the following methods to control for confounding:
 - i) Studies controlling for time-invariant unobservable confounding, including difference-in-differences, or fixed- or random-effects models with an interaction term between time and intervention for pre-intervention and post-intervention observations; and
 - ii) Studies assessing changes in trends in outcomes over a series of time points (interrupted time series, ITS), with or without contemporaneous comparison (controlled ITS), with sufficient observations to establish a trend and control for effects on outcomes due to factors other than the intervention (e.g. seasonality).
- d) Non-randomized studies with control for observable confounding, including non-parametric approaches (e.g. statistical matching, covariate matching, coarsened-exact matching, propensity score matching) and parametric approaches (e.g. propensity-weighted multiple regression analysis).

We will exclude all studies that do not fall under any of the criteria defined above. Examples of excluded study types are: simulation studies that aim to predict the effect of a certain intervention; observational studies with no control for selection bias; life-cycle analysis; process evaluations; acceptability studies; and non-systematic literature reviews.

f. Exclusion criteria

We will exclude any studies not meeting the criteria outlined in the above criteria a-e. That is, we will exclude studies with interventions that do not focus on sectors. We will exclude all studies that do not clearly articulate a comparison/control group (e.g. process evaluations). As indicated above, we will also exclude studies that do not focus on populations in Annex I countries or which do not report separate results for Annex I and non-Annex I countries. Studies published before the year 2000 will be excluded.

5. SEARCHING FOR EVIDENCE

a. Search steps

A comprehensive search strategy will be adopted to search research literature for qualifying studies to identify all available evidence that is relevant to the review question. The key objective of the strategy is to be sensitive rather than specific by deliberately formulating search strings and search sources that are over-inclusive. This may increase the number of citations to be screened but it reduces the risk of missing any relevant studies. The search strategy aims to find both academic and "grey" literature. A three-pronged search strategy will be employed in this review: (i) formal search of academic databases using pre-defined and explicit search strings and Boolean operators; (ii) a formal search of grey literature in key organizational websites using keywords but applying full search strings in cases where institutional databases allow the application of Boolean operators; and (iii) backward and forward citation searches of included and seminal studies. The full search strategy is available in Appendix 4.

b. Search databases and repositories

The database choice is guided by relevance and comprehensiveness in covering sectorial literature. We plan to conduct searches on the most appropriate databases for published literature and for grey literature, specialist organizational websites and research institutes, as highlighted in Table 3 below.

Table 3. List of targeted databases

DATABASE TYPE	NAME OF DATABASE
Academic	Centre for Agricultural Bioscience International (CABI) Abstracts
	PubMed
	Scopus
	Web of Science (Social Science Citation Index, Science Citation Index Expanded, Emerging Sources Citation Index via EBSCO
	Africa Wide Information
	Biological and Agricultural Index
	Business Source Ultimate
	EconLit
	GreenFILE
	Political science complete
	PsychInfo
	Urban studies abstracts
	Waters and Oceans Worldwide
	Supplementary searches ⁷
	AGRIS
Behavioural Public Policy	
Decision-A Journal for Research about Judgment and Decision Making	
Grey literature	African Development Bank (AfDB): https://www.afdb.org/en
	Asian Development Bank: https://www.adb.org/

⁷ We will carry out supplementary independent hand searches in two academic journals that are known to be hotspots of behavioural science. The journals are not covered by the bibliometric databases above but are identified as being particularly relevant.

DATABASE TYPE	NAME OF DATABASE
	Behaviour and Health Research Unit, University of Cambridge, UK: www.bhru.iph.cam.ac.uk/
	Behavioural Economics in Action at Rotman University of Toronto, CA: www.rotman.utoronto.ca/FacultyAndResearch/ResearchCentres/BEAR
	Behaviour Economics Team of the Australian Government, AUS: www.behaviouraleconomics.pmc.gov.au/
	Behavior Evidence hub: https://www.bhub.org/
	Behavioural Insights Team: https://www.bi.team/
	Behavioral Science and Policy Association: www.behavioralpolicy.org/
	Bill & Melinda Gates Foundation: https://www.gatesfoundation.org/
	Campbell Collaboration: https://campbellcollaboration.org/
	CEEDER: https://environmentalevidence.shinyapps.io/CEEDER/
	Center for Effective Global Action (CEGA) Research Publications: https://vcresearch.berkeley.edu/research-unit/center-effective-global-action
	Deloitte Insights: www2.deloitte.com/insights/us/en.html
	Department for Environment, Food & Rural Affairs, UK: https://www.gov.uk/government/organisations/department-for-environment-food-rural-affairs
	Environment Agency, UK: www.gov.uk/government/organisations/environment-agency
	Environmental Protection Agency, USA: www.epa.gov/
	Environmental Evidence Library: http://www.environmentalevidence.org/completed-reviews
	European Commission Joint Research Centre, EU: www.ec.europa.eu/jrc/en
	European Environment Agency, EU: www.eea.europa.eu/
	Federal Environment Agency, GER: www.umweltbundesamt.de/
	Federal Ministry for the Environment, Nature Conservation and Nuclear Safety, GER: www.bmu.de/
	Federal Ministry of Food and Agriculture, GER: www.bmel.de/
	Foreign, Commonwealth & Development Office: https://www.gov.uk/government/organisations/foreign-commonwealth-development-office
	Green Climate Fund: https://www.greenclimate.fund/publications

DATABASE TYPE	NAME OF DATABASE
	Green Finance Platform: https://www.greenfinanceplatform.org/
	Harvard Kennedy School Centre for Public Leadership, Behavioral Insights Group, US: www.cpl.hks.harvard.edu/behavioralinsights-group
	Ideas42: https://www.ideas42.org/
	Innovations for Poverty Action (IPA) Publications: https://www.poverty-action.org/publications
	Inter-American Development Bank: https://www.iadb.org/en/topics-effectiveness-improving-lives/impact-evaluations-repository
	International Fund for Agricultural Development: https://www.ifad.org/en/
	International Initiative for Impact Evaluation: 3ie Development Evidence Portal: https://developmentevidence.3ieimpact.org/
	International Institute for Environment and Development: www.iied.org/
	J-PAL: https://www.povertyactionlab.org/evaluations
	Millennium Challenge Corporation: https://www.mcc.gov/
	National Bureau of Economic Research: https://www.nber.org/
	NSW Government Behavioural Insights Unit, AUS: www.nsw.gov.au/behavioural-insights-unit
	Nudge Lebanon https://nudgelebanon.org/
	Observatory for Public Sector Innovation (OPSI) https://oecd-opsi.org/bi-projects/
	Organisation for Economic Co-operation and Development: http://www.oecd.org/
	PBL Netherlands Environmental Assessment Agency, NL: www.pbl.nl/en/
	Rare: www.rare.org
	The European Nudge Network: www.tenudge.eu/
	The Food and Agriculture Organization of the United Nations: www.fao.org/home/en/
	The London School of Economics and Political Sciences (LSE), Centre for Analysis of Risk and Regulation, UK: www.lse.ac.uk/accounting/CARR
	The World Bank: www.worldbank.org/
	Thünen-Institute, GER: www.thuenen.de/
	United Nations Development Programme (UNDP): www.undp.org/

DATABASE TYPE	NAME OF DATABASE
	United Nations Environment Programme (REDD+): https://www.unenvironment.org/explore-topics/climate-change/what-we-do/mitigation
	United Nations Framework Convention on Climate Change: https://unfccc.int/
	United Nations Food and Agriculture Organization (FAO): https://www.fao.org/home/en
	United States Department of Agriculture, USA: www.usda.gov/
	USAID Evaluations Clearinghouse: http://dec.usaid.gov/
	World Bank eLibrary: https://elibrary.worldbank.org/

c. Forward and backward citation searches

We will carry out backward citation searches by searching the reference lists of included studies, especially SRs and seminal papers and forward citation searches using Google Scholar to find papers that are cited in included studies.

d. Search terms

Our search terms will provide broad but manageable coverage related to the EGM and SR objective. We will design a series of sets of search terms with individual terms including wild card symbols (*) where appropriate, separated by the Boolean operator " AND ". The initial search terms are organized in the following categories.

- **Developing country terminology:** This sub-category includes terms that are often used interchangeably with or closely related to " income countries " including " under developed country names .
- **Methods terminology:** This category includes terminology related to the measurement and tracking of impacts such as " impact evaluation analysis " ; articulation of comparison groups related to the specific empirical methods such as " instrumental variables " . These do not always refer to explicit comparison groups but generate estimates of causal impacts that are comparative.
- **Intervention terminology:** Terms related to the behavioural science intervention areas of interest highlighted in the ToC above, as behavioural science intervention tools such as " nudge " , " choice architecture " . The development of the intervention terms is aimed to be broad and encapsulate numerous synonyms without limiting it to the technical definition of behavioural science. This will ensure that the search is wide enough not to miss relevant studies.
- **General restrictions:** This category is a combination of language and time-specific restrictions to enable us to restrict (academic database) search results to English-language articles and SRs published in peer-reviewed academic journals in or after the year 2000.

The search terms are subject to refinement in subsequent search trials. The robustness of our search will be tested using a list of benchmark papers, that is, studies that ought to be in the search results. In cases where two-thirds of the benchmark papers are retrieved through the database searches, the search strategy passes the robustness test.

e. Combination of search terms

The first substring is focused on the region of the PICOS elements of the research question. Synonyms for developing countries identified are combined using the OR Boolean operator. The second substring is on the methodology of studies of interest to the review, the " S " part of the PICOS. The third group of substrings is the intervention terms divided into five search substrings shown in Appendix 4, the " I " component of the PICOS. These behavioural science synonyms are used to improve the search. The overall combination of search concepts will follow the below syntax:

⁸ Available at <https://www.bhub.org/>

(1) L M I C ' P A N D I (2) m e t h o d A N D (' 3 S) ' I n t e r n e t r e m o v e m e n t o n ' I

f. Screening of studies

Review management software (EPPI-Reviewer 4) will be used to manage the entire review process. All potentially relevant citations gathered from the academic sources above will be imported into EPPI-Reviewer 4. They will undergo a manual screening process to be assessed for eligibility using the inclusion criteria highlighted above, and decisions made about each citation will be recorded on the same platform. Search results from organizational websites and the citation searches will be captured in MS Word, and only studies deemed to be relevant for the map will be transferred to EPPI-Reviewer 4. Studies that are not already on EPPI-Reviewer will be captured manually on the software. Before proceeding with screening, all duplicates of titles will be excluded from the review using the duplicate control function on EPPI-Reviewer 4.

To achieve both speed and quality in the screening process, we will be utilizing the machine learning algorithm function of EPPI-Reviewer 4, specifically the classifier. The classifier is a machine learning system of EPPI that allows the organization of studies into groups based on their probability of inclusion in the review. The bespoke classifier will be built after double screening a minimum of 10 per cent of all citations, enough from selected choices. To enhance the efficiency of the classifier, we will pay particular attention to the reasons for exclusion during the double-screening reconciliation process of the initial 10 per cent of citations. This then allows all studies with less than 20 per cent probability of inclusion to be automatically excluded from the review. We will then screen a random sample of the automatically excluded studies to double-check the accuracy of the function, and if all are excludable, we can auto-exclude the rest of the citations. A double-screening exercise at title and abstract will focus on all records with the likelihood of inclusion at 20 per cent or greater. Full-text documents of the remaining studies will also be screened by two reviewers. A third-party arbitrator will resolve any disagreements at both stages of the screening process. The screening process will be reported using a Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow chart. We will test reviewer bias (interrater reliability) at the start of each stage of the screening process using a Kappa analysis (Collaboration for Environmental Evidence, 2013). Two reviewers will screen a common random sample of 10 per cent of abstracts. The level of agreement between the number of articles rejected or accepted by the Kappa statistic will be calculated on a scale that ranges from 1 (perfect agreement) to -1 (strong disagreement). The individual screening will only be permissible once a Kappa statistic score of 0.85 or above is achieved.

6. DATA COLLECTION AND ANALYSIS

a. Data extraction and management

We will use a predefined data extraction tool to extract data systematically and transparently from the included primary studies and SRs. The coding framework tool highlighted in Appendix 5 will be translated into EPPI-Reviewer 4 to extract information that is required for both the evidence mapping and the in-depth review and synthesis.⁹ The data will be entered directly into the EPPI-Reviewer database and full-text reports will be examined and studies coded on variables related to:

- Descriptive data including authors, publication date and status, as well as other information to characterize the study including country, type of intervention, outcome, population, and context.

⁹ The data extraction for the EGM will focus on key descriptive variables in (a) whilst the data extraction for the SR, goes substantially beyond the descriptive details to incorporate aspects highlighted in (b).

- Information on intervention design, including implementation fidelity, how the intervention considers equity, participant uptake and adherence, contextual factors, and programme mechanisms.

To ensure consistency of coding quality, two reviewers will pilot the data extraction tool, working independently on a random sample (10 per cent) of eligible studies selected to test the tool on the complete range of the included impact evaluation designs and methods. The process will be repeated until a very high level of consistency – defined by a minimum Kappa statistic score of 0.85 – in the reviewer's application of codes is achieved. In the next stage, the remaining studies will be coded by individual reviewers, with a subset of these full texts being coded by different combinations of two reviewers independently extracting information from each study and then comparing their decisions. Any uncertainties or disagreements will be resolved via discussion to further review the study reports. An extra third-party arbitrator will resolve any outstanding disagreements.

b. Critical appraisal

We will apply a critical appraisal tool to assess the impact of bias on the trustworthiness of primary impact evaluations included in the SR.¹⁰ Trustworthiness refers to the confidence of the review team that the findings reported in the included studies used for the synthesis were rigorous and credible. To assess the risk of bias of the primary studies, we will adapt the Cochrane risk-of-bias tool for randomized and non-randomized studies (Sterne and others, 2016), which we have previously used and adapted in international development reviews (Stewart and others, 2015; Langer and others, 2017). Sterne and colleagues used a domain-based risk of bias tool covering the following six indications of trustworthiness: (i) selection bias; (ii) confounding bias; (iii) bias due to departures from applied interventions; (iv) bias due to missing data; (v) bias due to measurement of outcomes; and (vi) bias due to selection of the reported result. Each domain of bias will receive a low, moderate, high or critical risk of bias rating, allowing for a transparent calculation of the overall risk of bias score for each study. Studies with a critical risk of bias will be included in the review but excluded from the synthesis.

The critical appraisal tool used to assess studies for the SR is presented in Appendix 5. It will be piloted using a similar approach to that used for the piloting of the data extraction tool. Two reviewers will independently assess each study and then come together to compare their decisions. Where these reviewers are in disagreement about the risk of bias rating for a particular study, a third reviewer will be consulted.

c. Methods for handling dependent effect sizes

i. Criteria for the determination of independent findings

Complex data structures are a common occurrence in meta-analyses of impact evaluations. There are numerous scenarios through which these complex structures with dependent effect sizes might occur. For example, there could be several publications that stem from one study, or several studies based on the same data set. Some studies might have multiple treatment arms that are all compared to a single control group. Other studies may report outcome measurements from several time points or use multiple outcome measures to assess related outcome constructs. All such cases yield a set of statistically dependent effect size estimates (Borenstein and others, 2009).

¹⁰ In the EGM, we will only appraise included SRs for their trustworthiness. We intend to apply either the ROBINS (available at <http://www.bristol.ac.uk/population-health-sciences/projects/robis/>) or SURE (available at https://www.cardiff.ac.uk/data/assets/pdf_file/0007/1142962/SURE-CA-form-for-SR_2018.pdf) tools to appraise SRs. Impact evaluations included in the EGM will not be appraised in the EGM.

The research team will assess the extent to which relationships exist across the studies included in the review, and will avoid double counting of identical evidence by linking papers before data analysis. Where we have several publications reporting on the same effect, we will use effect sizes from the most recent publication. We will utilize information provided in studies to support these assessments, such as sample sizes, programme characteristics and key implementing and/or funding partners.

We will extract effects reported across different outcomes or subgroups within a study, and where information is collected on the same programme for different outcomes at the same or different periods, information on the full range of outcomes over time will be extracted. Where studies report effects from multiple model specifications, we will use the specification with the most controls. If this is not stated or is unclear, the specification with the most controls will be used. Where studies report multiple outcomes or evidence according to sub-groups of participants, we will record and report data on relevant sub-groups separately. Further information on criteria for determining independent effect sizes is presented below.

We will deal with dependent effect sizes through data processing and selection techniques, that utilize several criteria to select one effect estimate per study. When we have several publications reporting on the same study, we will use effect sizes from the most recent publication. For studies with outcome measures at different time points, we will follow De La Rue and colleagues (2014) and synthesize outcomes measured immediately after the intervention (defined as 1-6 months) and at follow-up (longer than six months) separately. If multiple time points exist within these periods, we adopt the most recent measure. We anticipate that many of the interventions that we will include in our review would be ongoing programmes and the follow-up would, therefore, reflect duration in a programme rather than time since the intervention. When such studies report outcome measures at different time points, we identify the most common follow-up period and include the follow-up measures that match this most closely in the meta-analysis. When studies include multiple outcome measures to assess related outcome constructs, we will follow Macdonald and others (2012) and select the outcome that appears to reflect the construct of interest most accurately without reference to the results.

If studies include multiple treatment arms with only one control group and the treatments represent separate treatment constructs, we calculate the effect size for treatment A versus control and treatment B versus control, and include them in separate meta-analyses according to the treatment construct. If treatments A and B represent variations of the same treatment construct, we calculate the weighted mean and standard deviation for treatment A and B before calculating the effect size for the merged group versus control group, following the procedures outlined by Borenstein and colleagues (2009, chapter 25). Where different studies report on the same programme but use different samples (e.g. from different regions, or separately for men and women) we included both estimates, treating them as independent samples, provided that effect sizes are measured relative to separate control or comparison groups.

ii. Effect size calculations

Quantitative data for outcome measures, including outcome descriptive information, sample size in each intervention group, outcomes means and standard deviations, and test statistics (e.g. t-test, F-test, p-values, 95 per cent confidence intervals) will be extracted using Excel. Effect size data will be stored, and any necessary cleaning will be conducted in Excel. Following the screening and descriptive data extraction process of ensuring consistency in coding quality, two reviewers will pilot the effect size data extraction tool, working independently on a random sample (10 per cent) of included studies to test the tool across a range of the included impact evaluation designs and methods. We aim to achieve a minimum Kappa statistic score of 0.90 following a round of repeating

the process for the tool to be finalized. After the piloting stage, the remaining studies will be coded by individual reviewers and all data extracted will be checked by a third reviewer.

An effect size expresses the magnitude (or strength) and direction of the relationship of interest (Valentine and others, 2015; Borenstein and others, 2009). We will extract data from each study to calculate standardized effect sizes for cross-study comparison wherever possible. For continuous outcomes comparing group means in a treatment and control group, we will calculate the standardized mean difference (SMD) using formulae provided in Borenstein and colleagues (2009). An SMD is a difference in means between the treatment and control groups divided by the pooled standard deviation of the outcome measure. Coherence can be biased in cases where sample sizes are small. Therefore, in all cases we adjust d using Hedges' method, Hedges uses the following formula (Ellis, 2010):

$$\left(1 - \frac{3}{4(n_1 + n_2) - 9}\right)$$

We choose an appropriate formula for effect size calculations in reference to, and dependent upon, the data provided in included studies. For example, for studies reporting means (X) and pooled standard deviation (SD) for treatment (T) and control or comparison (C) at follow up only:

$$= \frac{X_{+1} - X_{-1}}{SD_{+1}}$$

If the study does not report the pooled standard deviation, it is possible to calculate it using the following formula:

$$SD_{+1} = \frac{\sqrt{(X_{+1} - 1)^2_{+1} + (X_{+1} - 1)^2_{-1}}}{\sqrt{n_{+1} + n_{-1} - 2}}$$

Where the intervention is expected to change the standard deviation of the outcome variable, we use the standard deviation of the control group only.

For studies reporting means (\bar{X}) and standard deviations (SD) for treatment and control or comparison groups at baseline (p) and follow up (p+1):

$$= \frac{\bar{X}_{+1} - \bar{X}_{-1}}{SD_{+1}}$$

For studies reporting mean differences (\bar{D}) between treatment and control and standard deviation (SD) at follow up (p+1):

$$= \frac{\bar{D}_{+1}}{SD_{+1}} = \frac{\bar{D}_{+1} - \bar{D}_{-1}}{SD_{+1}}$$

For studies reporting mean differences between treatment and control, standard error (SE) and sample size (n):

$$= \frac{\bar{D}_{+1}}{SE_{+1}}$$

As primary studies have become increasingly complex, it has become commonplace for authors to extract partial effect sizes (e.g. a regression coefficient adjusted for covariates) in the context of meta-analysis. For studies reporting regression results, we will follow the approach suggested by (Keef and Roberts, 2004) using the regression coefficient and the pooled standard deviation of the outcome. Where the pooled standard deviation of the outcome is unavailable, we utilize regression coefficients and standard errors or t-statistics to do the following, where sample size information is available in each group:

$$= \frac{1}{n} + \frac{1}{n}$$

where n denotes the sample size of the treatment group and control.

We will use the following where only the total sample size information (N) is available, as suggested in (Polanin and others, 2016):

$$= \frac{2}{N} = \frac{4}{N} + \frac{2}{N}$$

We calculate the t -statistic (t) by dividing the coefficient by the standard error. If the authors only report confidence intervals and no standard error, we calculate the standard error from the confidence intervals. If the study does not report the standard error but reports t , we extract and use this as reported by the authors. In cases in which significance levels are reported rather than t or SE (b), then t was imputed as follows:

$$\begin{aligned} \text{Prob} > 0.1: t &= 0.5 \\ 0.1 < \text{Prob} < 0.05: t &= 0.8 \\ 0.05 < \text{Prob} < 0.01: t &= 1.4 \\ 0.01 < \text{Prob} < 0.001: t &= 2.8 \end{aligned}$$

Where outcomes are reported in proportions of individuals, we calculate the Cox-transformed log odds ratio effect size (Sánchez-Meca and others, 2003):

$$= \frac{(\ln OR)}{1.65}$$

where OR is the odds ratio calculated from the two-by-two frequency table.

Where outcomes were reported based on proportions of events or days, we use the standardized proportion difference effect size:

$$= \frac{p_t - p_c}{p}$$

where p_t is the proportion in the treatment group and p_c the proportion in the comparison group, and the denominator is given by:

$$p = \frac{p_t + p_c}{2}$$

where p is the weighted average of p_c and p_t :

$$= \frac{p_t + p_c}{2}$$

An independent reviewer will evaluate a random selection of 10 per cent of effect sizes to ensure that the correct formulae will be employed in effect size calculations. In all cases after synthesis, we will convert the pooled effect sizes to commonly used metrics such as percentage changes and mean differences in outcome metrics typically used (e.g. weight in kg) whenever feasible.

d. Visualization of the evidence gap map

We will use Africa Centre for Evidence's (ACE) the EGM. Appendix 2B shows the respective visualization based on the intervention-outcome matrix structure in Appendix 2A to map the identified evidence-base consisting of different behaviour intervention categories mapped across the intermediate, final, development results and impact outcomes. Following the data extraction process in EPPI-Reviewer, we will generate and export a JavaScript Object Notation (JSON) for EGM. The "design" function of the mapping software preferences, select colour codes that will allow the visualization of included studies to be separated

by pre-defined characteristics. For example, studies can be separated by study type, that is, impact evaluations (green-coloured bubbles) and SRs (yellow-coloured bubbles) as highlighted in Appendix 2B. The software options also allow users to tailor the evidence-base to their own contexts using filters (e.g. sector, region, study design).

e. Data synthesis

From selected quantitative studies in the -EGM's analyses of studies that are assessed to be sufficiently similar, and only combine studies using meta-analysis when we identify two or more effect sizes using a similar outcome construct and where the comparison group state is judged to be similar across the two (c.f. the approach taken by Wilson and colleagues, 2011). We will combine studies in the same analysis when they evaluate the same intervention type and the same outcome type. Where there were too few studies or included studies are considered too heterogeneous in terms of interventions or outcomes, we will discuss the individual effect sizes along the causal chain. As programme theory of interventions suggests that there will be heterogeneity across studies, we will adopt inverse-variance weighted, random effects meta-analytic models (Higgins and others, 2020) to account for this.

We will conduct separate analyses for the major outcome categories for each intervention type: intermediate outcomes, final outcomes, development results and impact outcomes that are identified. Based on an analysis of the interventions that we find, we attempt to further elaborate on the pathway of change that was outlined above to the extent possible. We aim to conduct moderator analyses to try to explain variations in effect sizes. Moderators are variables such as socio-economic context and population characteristics, measured at baseline, that interact with treatment to change outcome for each group (Pincus and others, 2011). Moderator analyses will be reported in a tabular format below each meta-analysis, calculated using the same one-way random effects ANOVA model as applied to the moderator analyses.

f. Assessment of heterogeneity

To visibly examine variability in the effect size estimates, we will use forest plots to display the estimated effect sizes from each study along with their 95 per cent confidence intervals. Subsequently, and acknowledging the limitations of quantification of heterogeneity and the different strengths of statistical approaches, the following test for heterogeneity will be conducted: calculation of the Q- statistic as a statistical test of heterogeneity (Hedges & Olkin, 1985); and calculation of the i^2 and Tau^2 statistic to provide estimates of the magnitude of the variability across study findings caused by heterogeneity (Higgins and Thompson, 2002; Higgins and others, 2003; Borenstein and others, 2009).

g. Sensitivity analyses

To test the robustness of the results of the meta-analysis, a number of sensitivity analyses will be conducted. Broadly, this involves collecting data on and assessing the sensitivity of findings to (i) the methods of the primary studies and (ii) the methods of the review. We anticipate that the included studies will vary methodologically and therefore conduct sensitivity analyses to examine the influence of these variations on the summary measures, in order to offer possible explanations for the differences between studies when interpreting the results. We will examine whether the results were sensitive to study design, the risk of bias associated with the study, the degree of missing/incomplete data, and the way outcomes are measured and the timing at which they are measured. The main objective of the sensitivity analysis is to serve as a visual tool that allows informal comparisons to determine whether the results of our meta-analyses are sensitive to the

methodological decisions of the review team. The sensitivity analyses will be carried out by adopting a one-way random-effects ANOVA model calculated in EPPI-Reviewer.

h. Strength of the evidence assessment

The last research step in the SR will be to conduct a Grading of Recommendations, Assessment, Development and Evaluations (GRADE) assessment to report on the overall strength of the evidence base and recommendations made based on the synthesis of the review. This step is distinct from the critical appraisal step and considers additional factors to assess the overall body of the evidence and how much trust can be placed in recommendations that are made based upon it. Appendix 7 presents the GRADE tool with hypothetical decisions for illustration purposes.

D. DISTRIBUTION OF MAIN ROLES

PROJECT FUNDING, OVERSIGHT AND CO-CREATION	
Dr. Martin Prowse	Independent Evaluation Unit, Green Climate Fund, Task manager
Deborah Sun Kim	Independent Evaluation Unit, Green Climate Fund
Andreas Reumann	Independent Evaluation Unit, Green Climate Fund
Prof. Dr. Jyotsna Puri	International Fund for Agricultural Development
Dr. Romina Cavatassi	International Fund for Agricultural Development
PROJECT EXECUTION	
Africa Centre for Evidence	
Dr. Laurenz Langer	Project oversight and management Co-PI: Project oversight and management, stakeholder and client engagement, finance and reporting, drafting and finalization of outputs and deliverables. Research: Mapping and synthesis Synthesis method lead: design of all research activities, tools development, and research staff training and support where relevant. Lead on formulating inclusion criteria, EGM framework development, meta-analysis, GRADE assessment, and qualitative comparative analysis if conducted
Dr. Carina van Rooyen	Project oversight and management Co-PI: Project oversight and management, stakeholder and client engagement, finance and reporting, drafting and finalization of outputs and deliverables. Research: Mapping and synthesis Advice on adaptation and contextualization of synthesis methodologies in the environmental sector
Prof. Ruth Stewart	Synthesis advisor: Ad hoc advice and technical inputs related to complex evidence synthesis, climate change and behavioural science issues encountered
Promise Nduku	Research lead Research: Mapping and synthesis Synthesis specialist: design and conduct of search strategy, screening, data extraction, and critical appraisal for both the EGM and the SR. Lead on EGM visualization and effect size calculation
Andile Madonsela	Research: Mapping and synthesis Research assistant: research support in accessing full-text articles, cataloguing data, collating background information, and editorial support

Shona Putuka	Research: Mapping and synthesis Research assistant: research support in accessing full-text articles, cataloguing data, collating background information, and editorial support.
--------------	---

Content experts

Dr. Benjamin Curtis	Subject input related to behavioural science on the following areas: scope of the EGM and SR including ToC; inclusion criteria for the EGM and SR; framework development for the EGM; reviewing search strategy and output; data extraction variables for SR; interpreting synthesis results; output review including approach paper, EGM report and SR protocol, and technical report; and stakeholder and client engagement.
---------------------	--

Dr. Caitlin Blaser Mapitsa	Content advisor: Ad hoc advice and technical inputs related to complex evidence synthesis, ToC development, climate change, and behavioural science issues encountered.
----------------------------	---

Jamie Robertsen	Subject input related to climate change on the following areas: scope of the EGM and SR including ToC; inclusion criteria for the EGM and SR; framework development for the EGM; reviewing search strategy and output; data extraction variables for SR; interpreting synthesis results; output review including approach paper, EGM report and SR protocol
-----------------	---

Samantha Booth	Content advisor: Ad hoc advice and technical inputs related to complex evidence synthesis, ToC development, climate change, and behavioural science issues encountered
----------------	--

REFERENCES

- Balmford, A., and others (2021). Making more effective use of human behavioural science in conservation interventions. *Biological Conservation*, 261, 109256.
- Binet, S., and others (2021). *Independent evaluation of the adaptation portfolio and approach of the Green Climate Fund*. Independent Evaluation Unit, Green Climate Fund. Songdo, South Korea.
- Bollinger, Bryan, Kenneth. T. Gillingham, and M. Ovaere (2020). Field experimental evidence shows that self-interest attracts more sunlight. *Proceedings of the National Academy of Sciences*, 117(34), 20503-20510.
- Borenstein, M., and others (2009). *Introduction to Meta-Analysis*. Chichester, West Sussex: John Wiley and Sons.
- Bours, D., C. McGinn, and P. Pringle (2014). Guidance note 3: Theory of Change approach to climate change adaptation programming. Available at http://www.managingforimpact.org/sites/default/files/resource/2014_02_sea_change_ukcip_gn_3_toc_approach_to_cca_programming.pdf.
- Byerly, H., and others (2018). Nudging behaviour: evidence and opportunities. *Frontiers in Ecology and the Environment*, 16(3), 159-168.
- Cheng, T., D. K. Woon, and J. K. Lynes (2011). The use of message framing in the promotion of environmentally sustainable behaviors. *Social Marketing Quarterly*, 17(2), 48-62.
- Cinner, J. (2018). How behavioral science can help conservation. *Science*, 362(6417), 889-890.
- Collaboration for Environmental Evidence (2013). Guidelines for Systematic Review and Evidence Synthesis in Environmental Management. Version 4.2. Available at www.environmentalevidence.org/Documents/Guidelines/Guidelines4.2.pdf.
- Datta, S., and S. Mullainathan (2014). Behavioral design: a new approach to development policy. *Review of Income and Wealth*, 60(1), 7-35. Available at <https://doi.org/10.1111/roiw.12093>.
- De La Rue, L., and others (2014). School-based interventions to reduce dating and sexual violence: A systematic review. *Campbell Systematic Reviews*, 10(1), 1-110. Available at <https://doi.org/10.4073/csr.2014.7>.
- Dhar, R. (2014). Why choosing healthy foods is hard, and how to help: Presenting the 4Ps framework for behaviour change. *Customer Needs and Solutions*, 253-262.
- Duflo, E., M. Kremer, and J. Robinson (2011). Nudging farmers to use fertilizer: Theory and experimental evidence from Kenya. *American Economic Review*, 101(6), 2350-90.
- Flanagan, A.E., and J. C. Tanner (2016). Evaluating behaviour change in international development operations: A new framework. World Bank IEG Working Papers 2016/02. Washington, D.C.: World Bank.
- Gifford, R., C. Kormos, and A. McIntyre (2011). Behavioral dimensions of climate change: Drivers, responses, barriers, and interventions. *Wiley Interdisciplinary Reviews: Climate Change*, 2(6), 801-827. Available at <https://doi.org/10.1002/wcc.143>.
- Global Commission on Adaptation (2019). Adapt now: A global call for leadership on climate resilience. Global Commission on Adaptation. Available at <https://gca.org/reports/adapt-now-a-global-call-for-leadership-on-climate-resilience/>.
- Hyland-Wood, B., and others (2021). Toward effective government communication strategies in the era of COVID-19. *Humanities and Social Sciences Communications*, 8(1), 1-11.
- Hallsworth, M., and others (2017). The behavioralist as tax collector: Using natural field experiments to enhance tax compliance. *Journal of Public Economics*, 148, 14-31. Available at <https://doi.org/10.1016/j.jpubeco.2017.02.003>.
- Hedges, L., and I. Olkin (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press. Available at <https://doi.org/10.1016/C2009-0-03396-0>.
- Higgins, J.P.T., and S. G. Thompson (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, 21, 1539-1558. Available at <https://doi.org/10.1002/sim.1186>.

- Kahan, D. M., and others (2012). The polarizing impact of science literacy and numeracy on perceived climate change risks. *Nature climate change*, 2(10), 732-735. Available at <http://doi.org/10.1038/NCLIMATE1547>.
- Krüger, C., and Puri, J. (2020). Going the last mile: Behaviour science and investments in climate change mitigation and adaptation. Available at https://ieu.greenclimate.fund/sites/default/files/document/201123-going-last-mile-top_0.pdf.
- Kollmuss, A., and Agyeman, J. (2002). Mind the gap: why do people act environmentally and what are the barriers to pro-environmental behavior? *Environmental Education Research*, 8(3), 239-260. Available at <https://doi.org/10.1080/13504620220145401>.
- Kusmanoff, A. M., and others (2020). Five lessons to guide more effective biodiversity conservation message framing. *Conservation Biology*, 34(5), 1131-1141. Available at <https://pubmed.ncbi.nlm.nih.gov/32043648/>.
- Macdonald, G., and M. Taylor (2012). Cognitive-behavioural interventions for children who have been sexually abused: A systematic review. *Campbell Systematic Reviews*, 8(1), 1-111. Available at <https://doi.org/10.4073/csr.2012.14>.
- Nisa, C. F., and others (2019). Meta-analysis of randomised controlled trials testing behavioural interventions to promote household action on climate change. *Nature communications*, 10(1), 1-13. Available at <https://doi.org/10.6084/m9.figshare.9641999>.
- Osbaldiston, R., and J. P. Schott (2012). Environmental sustainability and behavioral science: Meta-analysis of proenvironmental behavior experiments. *Environment and Behavior*, 44(2), 257-299. Available at <https://doi.org/10.1177/0013916511402673>.
- Pincus, T., and others J. (2011). Methodological criteria for the assessment of moderators in systematic reviews of randomised controlled trials: a consensus study. *BMC medical research methodology*, 11(1), 1-14. Available at <https://doi.org/10.1186/1471-2288-11-14>.
- Polanin, J.R., E. E. Tanner-Smith, and Hennessy, E.A. (2016). Estimating the difference between published and unpublished effect sizes: A meta-review. *Review of Educational Research*, 86(1), 207-236. Available at <https://doi.org/10.3102/0034654315582067>.
- Saunders, C. D., A. T. Brook, and O. L. I. N. Eugene Myers Jr (2006). Using psychology to save biodiversity. *Conservation Biology*, 20(2), 702-705. Available at <https://doi.org/10.1111/j.1523-1739.2006.00435.x>.
- Schot, J., L. Kanger, and G. Verbong (2016). The roles of users in shaping transitions to new energy systems. *Nature Energy*, 1(5), 1-7. Available at <https://doi.org/10.1038/nenergy.2016.54>.
- Schmuck, P., and W. P. Schultz (eds.) (2012). *Psychology of sustainable development*. Springer Science & Business Media. Available at <https://doi.org/10.1007/978-1-4615-0995-0>.
- Schultz, P. W. (2011). Conservation means behavior. *Conservation Biology*, 25(6), 1080-1083. Available at <https://doi.org/10.1111/j.1523-1739.2011.01766.x>.
- Schubert, C. (2017). Green nudges: Do they work? Are they ethical? *Ecological Economics*, 132, 329-342. Available at <https://doi.org/10.1016/j.ecolecon.2016.11.009>.
- Service, O., and others (2014). *EAST: Four simple ways to apply behavioural insights*. The Behavioural Insights Team.
- Siegel, L. (2012). *Loss and damage: Theme of slow onset events*. German Watch, Climate Development Network. Available at <https://germanwatch.org/sites/default/files/publication/6674.pdf>.
- Snilsveit, B., and others (2019). Incentives for climate mitigation in the land use sector—the effects of payment for environmental services on environmental and socio-economic outcomes in low and middle income countries: A mixed methods study. *Campbell Systematic Reviews*, 15(3), e1045. Available at <https://doi.org/10.1002/cl2.1045>.
- Solomon, S., and others (2007). *Climate change 2007: The physical science basis: Working Group I contribution to the fourth assessment report of the IPCC (Vol. 4)*. Cambridge University Press. Available at <https://www.ipcc.ch/report/ar4/wg1/>.

Stern, P. C., and others (2016). Opportunities and insights for reducing fossil fuel consumption by households and organizations. *Nature Energy*, 1(5), 1-6. Available at <https://doi.org/10.1038/nenergy.2016.43>.

Stoknes, P. E. (2014). Rethinking climate communication. *Energy Research & Social Science*, 1, 161-170.

Szaszi, B., and others (2018). A systematic scoping review of the choice architecture movement: Toward understanding when and why nudges work. *Journal of Behavioral Decision Making*, 31(3), 355-366.

Valentine, J.C., A. M. Aloe, and T. S. Lau (2015). Life after NHST: How to describe your data with integrity. *Basic and Applied Social Psychology*, 37(5), 260-273. Available at <https://doi.org/10.1080/01973533.2015.1060240>.

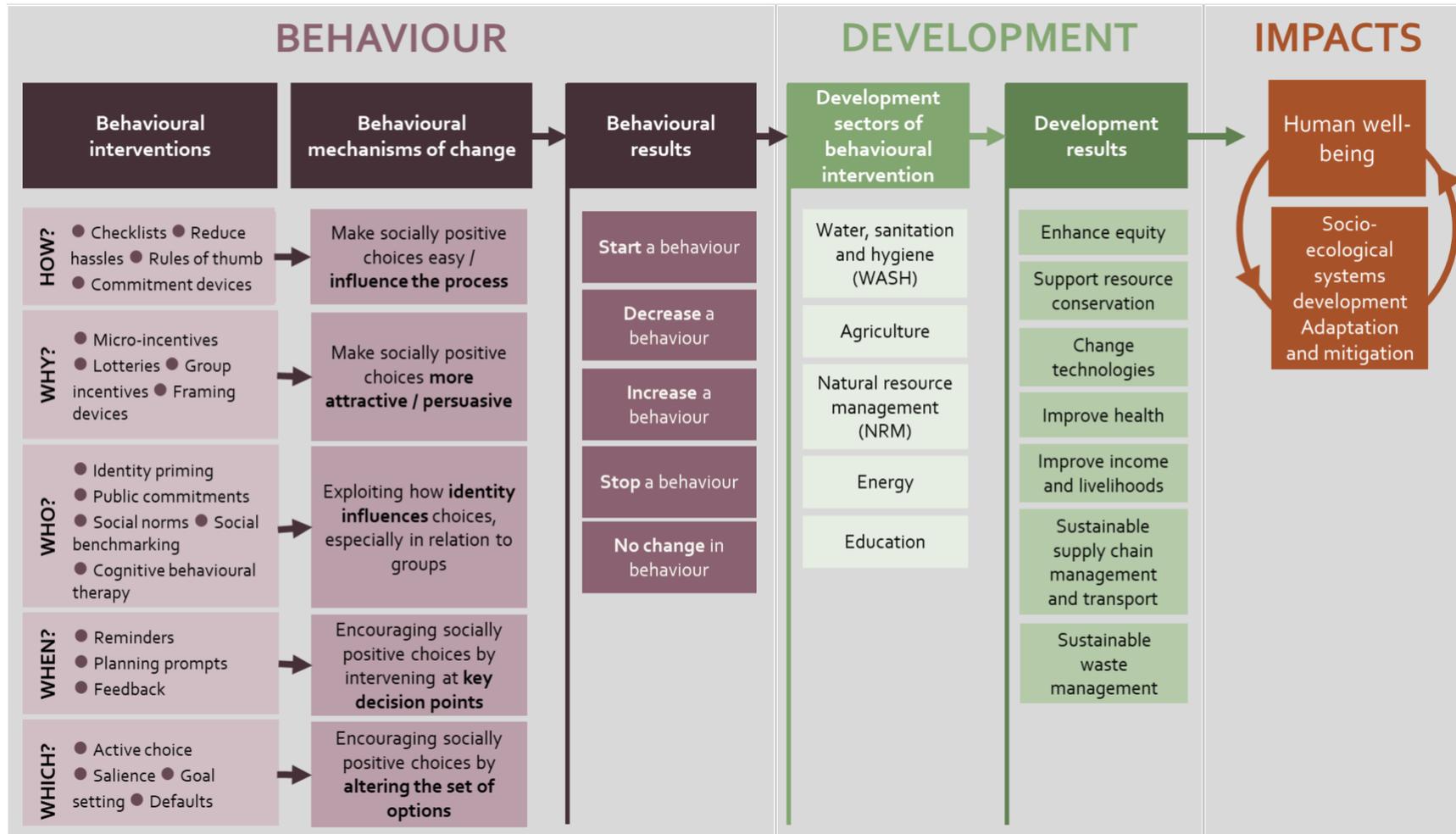
Van der Linden, S., E. Maibach, and Leiserowitz, A. (2015). Improving public engagement with climate change: Five "best practices" perspectives on insight. *Psychological Science*, 10(6), 758-763.

Wade, K., and M. Jennings (2015). How climate change will impact developed and developing countries. Climate change and the global economy: regional effects [online]. Available at <https://www.schroders.com/en/ch/asset-management/insights/economics/climate-change-and-the-global-economy-regional-effects/>. Accessed on 14 November 2021.

Wilson, D.B., D. Weisburd, and D. McClure (2011). Use of DNA testing in police investigative work for increasing offender identification, arrest, conviction and case clearance. *Campbell Systematic Reviews*, 7(1), 1-53. Available at <https://doi.org/10.4073/csr.2011>.

APPENDICES

Appendix 1. THEORY OF CHANGE (TOC)



Appendix 2. INTERVENTION-OUTCOME FRAMEWORK

A. INTERVENTION-OUTCOME FRAMEWORK MATRIX

INTERVENTIONS		OUTCOMES																	
		Intermediate outcomes (knowledge & attitudes)				Final outcome (behaviour change)					Development results						Socio-ecological systems development (includes human wellbeing)		
		Know of intervention	Take part in intervention	Acquire knowledge	Change attitudes	Start behaviour	Increase behaviour	Decrease behaviour	End behaviour	No change in behaviour	Enhance equity	Support resource conservation	Changing technologies	Improve health	Improve income and livelihoods	Sustainable supply chain management and transport	Sustainable waste management	Mitigation	Adaptation
How?	Checklists																		
Rules of thumb	Reduce hassles																		
Commitment devices																			

		OUTCOMES																	
		Intermediate outcomes (knowledge & attitudes)				Final outcome (behaviour change)					Development results						Socio-ecological systems development (includes human wellbeing)		
		Know of intervention	Take part in intervention	Acquire knowledge	Change attitudes	Start behaviour	Increase behaviour	Decrease behaviour	End behaviour	No change in behaviour	Enhance equity	Support resource conservation	Changing technologies	Improve health	Improve income and livelihoods	Sustainable supply chain management and transport	Sustainable waste management	Mitigation	Adaptation
Why?		Micro-incentives	Group incentives	Lotteries	Framing devices														

		OUTCOMES																			
		Intermediate outcomes (knowledge & attitudes)				Final outcome (behaviour change)					Development results						Socio-ecological systems development (includes human wellbeing)				
		Know of intervention	Take part in intervention	Acquire knowledge	Change attitudes	Start behaviour	Increase behaviour	Decrease behaviour	End behaviour	No change in behaviour	Enhance equity	Support resource conservation	Changing technologies	Improve health	Improve income and livelihoods	Sustainable supply chain management and transport	Sustainable waste management	Mitigation	Adaptation		
Who?		Social benchmarking	Social norms	Public commitments	Identity priming																

		OUTCOMES																	
		Intermediate outcomes (knowledge & attitudes)				Final outcome (behaviour change)					Development results						Socio-ecological systems development (includes human wellbeing)		
		Know of intervention	Take part in intervention	Acquire knowledge	Change attitudes	Start behaviour	Increase behaviour	Decrease behaviour	End behaviour	No change in behaviour	Enhance equity	Support resource conservation	Changing technologies	Improve health	Improve income and livelihoods	Sustainable supply chain management and transport	Sustainable waste management	Mitigation	Adaptation
When?		Feedback	Planning prompts	Reminders	Cognitive behavioural therapy														

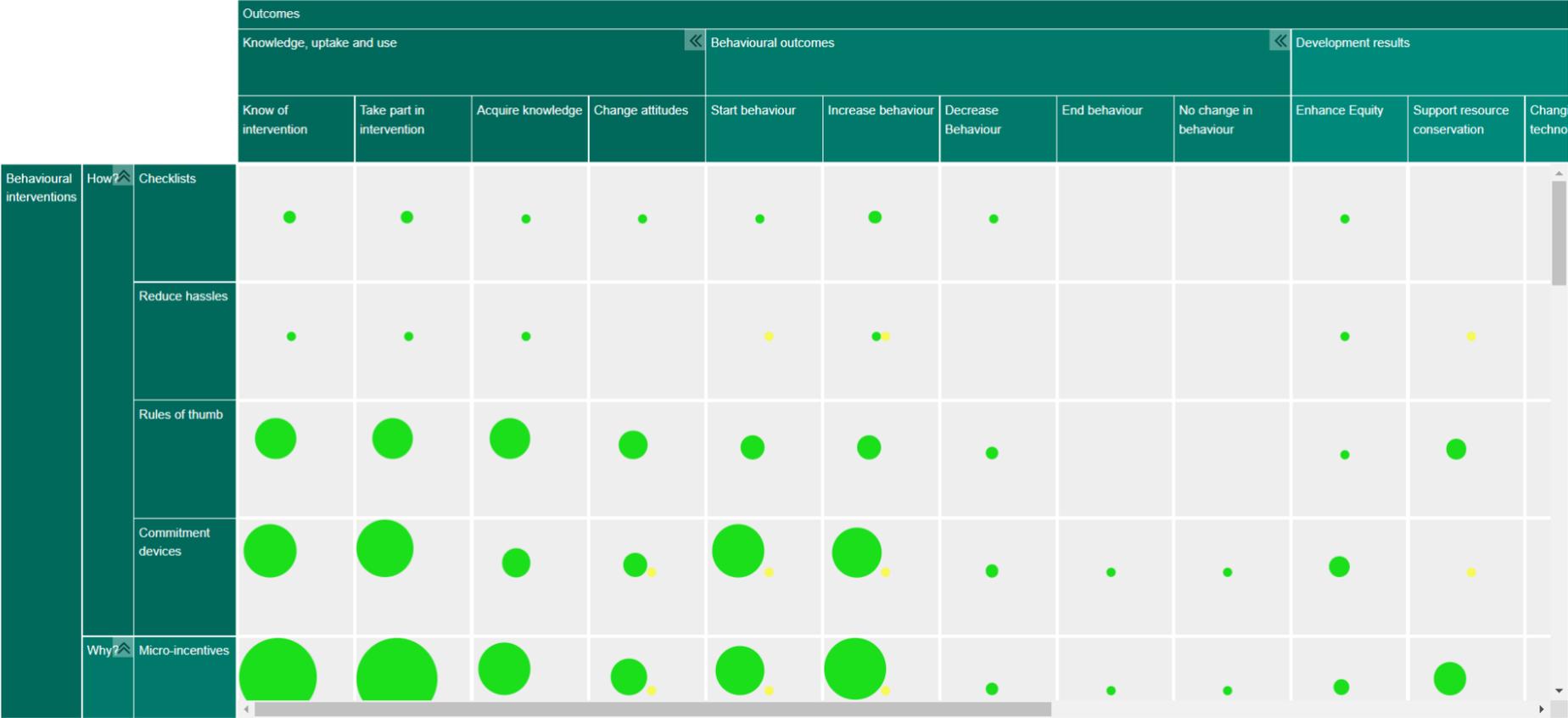
		OUTCOMES																
		Intermediate outcomes (knowledge & attitudes)				Final outcome (behaviour change)					Development results						Socio-ecological systems development (includes human wellbeing)	
		Know of intervention	Take part in intervention	Acquire knowledge	Change attitudes	Start behaviour	Increase behaviour	Decrease behaviour	End behaviour	No change in behaviour	Enhance equity	Support resource conservation	Changing technologies	Improve health	Improve income and livelihoods	Sustainable supply chain management and transport	Sustainable waste management	Mitigation
Goal setting																		
Defaults																		

B. EVIDENCE GAP MAP VISUALIZATION

Filters Hide Headers Exit Fullscreen About Submit a Study View Records



Behavioural science interventions within the human development and environmental fields in developing countries: An evidence gap map



● Impact evaluation ● Systematic review

Generated using v.2.1.0 of the EPPI-Mapper powered by EPPI Reviewer and created with ❤️ by the Digital Solution Foundry team.

Appendix 3. SUMMARY INCLUSION AND EXCLUSION CRITERIA ORGANIZED USING THE PICOS (POPULATION, INTERVENTION, OUTCOME, COMPARISON, STUDY DESIGN) MODEL

The tables below present a summary of our inclusion criteria for the EGM and the SR. They are intended for illustration and do not present an exhaustive outline of the inclusion criteria.

A. EVIDENCE GAP MAP

INCLUDED	INCLUSION DESCRIPTION	EXCLUSION DESCRIPTION
Population	<ol style="list-style-type: none"> 1) Studies that focus on behavioural science interventions in: <ol style="list-style-type: none"> a) Non-Annex 1 countries b) For primary studies non-Annex 1, and Annex 1 countries (jointly) if analyses distinguish effects across the two samples c) SRs are included in the EGM either if data is aggregated for non-Annex 1 countries relative to Annex 1 or if there is at least a single primary study included that is from non-Annex 1 countries 2) English-language literature 3) Publication date: 2000 onwards 	<ol style="list-style-type: none"> 1) Studies that focus on behavioural science interventions in: <ol style="list-style-type: none"> a) Annex 1 countries only for both primary studies and SRs b) Primary studies with a combination of both non-Annex 1 and Annex 1 countries if analyses does not distinguish the two samples 2) Non-English language literature 3) Studies published before the year 2000

INCLUDED	INCLUSION DESCRIPTION	EXCLUSION DESCRIPTION
Interventions	<p>Bi-sectorial focus on the environmental sector and human development sector.</p> <ul style="list-style-type: none"> a) Delivered at any administrative level b) Administered to any type of beneficiary (e.g. individual, household) c) By any type of actor (e.g. government, non-governmental organization) <p>These interventions include the following behavioural tools: checklists; reduce hassles; rules of thumb; commitment devices; micro-incentives; group incentives; lotteries; framing devices; identity priming; public commitments; social norms; social benchmarking; cognitive behavioural therapy; reminders; planning prompts; feedback; active choice; salience (communication); salience (experience design); goal setting; defaults</p> <p>Studies looking at behavioural science interventions with different:</p> <ul style="list-style-type: none"> a) modes of delivery; doses; durations; intensities; co-interventions b) degree of complexity; sample sizes 	<ul style="list-style-type: none"> 1) Interventions not in the environmental or human development sectors 2) Interventions focusing on trainings, capacity building initiatives, or farmer field schools
Comparator	<p>Studies that identify a comparison/control group</p>	<ul style="list-style-type: none"> 1) Descriptive/predictive analyses without a clear comparison/control group 2) Methods that do not utilize comparison/control groups (e.g. life-cycle assessment)
Outcomes	<p>Outcomes measured at a reasonable time after the onset of intervention following the behavioural science intervention leading to changes in intermediate outcomes (change in attitudes), final outcomes (behaviour change), development-related outcomes or socio-ecological systems development outcomes. A range of outcomes measured at the individual, household, community, and company level.</p> <p>Outcomes are organized into the following categories and sub-categories:</p> <ul style="list-style-type: none"> 1) Intermediate outcomes Know of intervention; take part in intervention; acquire knowledge; change in attitudes 2) Final outcomes Start behaviour; increase behaviour; decrease behaviour; end behaviour; no change in behaviour 3) Development results Enhance equity; support resource conservation; changing technologies; improve health; improve income and livelihoods; sustainable waste management; sustainable supply chain management and transport 	<ul style="list-style-type: none"> 1) Any outcomes not meeting the stated criteria

INCLUDED	INCLUSION DESCRIPTION	EXCLUSION DESCRIPTION
	4) Impact <ul style="list-style-type: none"> a) Socio-ecological systems development b) Mitigation; adaptation 	
Study design	1) Impact evaluations (experimental, quasi-experimental). For example: <ul style="list-style-type: none"> a) Randomized controlled trials b) Difference-in-differences design c) Regression discontinuity design d) Instrumental variable design e) Propensity score matching designs 2) Systematic reviews	1) Non-counterfactual impact evaluation designs 2) Non-systematic literature review

B. SYSTEMATIC REVIEW

INCLUDED	INCLUSION DESCRIPTION	EXCLUDED
Population	1) Studies that focus on behavioural science interventions in: <ul style="list-style-type: none"> a) Non-Annex 1 countries b) Non-Annex Annex 1, and Annex 1 countries (jointly) if analyses distinguish effects across the two samples 2) English-language literature 3) Publication date: 2000 onwards	1) Studies that focus on behavioural science interventions in: <ul style="list-style-type: none"> a) Annex 1 countries only b) Combination of both non-Annex I and Annex I countries if analyses do not distinguish the two samples 2) Non-English language literature 3) Studies published before the year 2000
Interventions	Bi-sectorial focus on the environmental sector and human development sector. <ul style="list-style-type: none"> a) Delivered at any administrative level b) Administered to any type of beneficiary (e.g. individual, household) c) By any type of actor (e.g. government, non-governmental organization) These interventions include the following behavioural tools: checklists; reduce hassles; rules of thumb; commitment devices; micro-incentives; group incentives; lotteries; framing devices; identity priming; public commitments; social norms; social benchmarking; cognitive behavioural therapy; reminders; planning prompts; feedback; active choice; salience (communication); salience (experience design); goal setting; defaults Studies looking at behavioural science interventions with different: <ul style="list-style-type: none"> d) modes of delivery; doses; durations; intensities; co-interventions e) degree of complexity; sample sizes 	1) Intervention not in the environmental sector and human development sector 2) Interventions focusing on trainings, capacity building initiatives, or farmer field schools
Comparator	Studies that identify a comparison/control group	1) Descriptive/predictive analyses without a clear comparison/control group 2) Methods that do not utilize comparison/control groups (e.g. life-cycle assessment)
Outcomes	Outcomes measured at any reasonable time after the onset of intervention following the behavioural since intervention, leading to changes in intermediate outcomes (change in attitudes), final outcomes (behaviour change), development-related outcomes or socio-ecological systems development.	1) Any outcomes not meeting the stated criteria

INCLUDED	INCLUSION DESCRIPTION	EXCLUDED
	<p>A range of outcomes measured at the individual, household, community, and company level Outcomes are organized into the following categories and sub-categories:</p> <ol style="list-style-type: none"> 1) Intermediate outcomes Know of intervention; take part in intervention; acquire knowledge; change attitudes 2) Final outcomes Start behaviour; increase behaviour; decrease behaviour; end behaviour; no change in behaviour 3) Development results Enhance equity; support resource conservation; changing technologies; improve health; improve income and livelihoods; sustainable waste management; sustainable supply chain management and transport 4) Impact <ol style="list-style-type: none"> a) Socio-ecological systems development b) Mitigation; adaptation 	
Study design	<p>Impact evaluations (experimental, quasi-experimental). For example:</p> <ol style="list-style-type: none"> a) Randomized controlled trials b) Difference-in-differences design c) Regression discontinuity design d) Instrumental variable design e) Propensity score matching designs 	<ol style="list-style-type: none"> 1) Non-counterfactual impact evaluation designs 2) Systematic reviews

" under developed and developing countries " OR " under de
" lower middle income countries " OR " lower
nation " OR " income countries " OR " upper - middle
income countries " OR " low income countries " OR " low
OR " lower income countries " OR " lower income
income nation and Middle East countries " OR " Low and M
OR " underserved country " OR " underserved country
nations " OR " underserved world " OR " under serv
" under served nation " OR " under served nation
OR " deprived countries " OR " deprived nation " O
" poor country " OR " poor countries " OR " poor na
" poorer country " OR " poorer countries " OR " poor
world " OR " developing economy " OR " developing
" less developed economies " OR " lesser develope
" under developed economies " OR " under developed o
" underdeveloped economies " OR " middle income e
" low income economy " OR " low income economies "
income economies " OR " transitional country " OR
OR " transitional country " OR " transitional cou
UMICs OR UMIC OR (" khmer " AND " republic ") OR (
" african " AND " republic ")

B. METHODOLOGY

" Systematic review " OR " longitudinal study " O
" comparison study " OR " Longitudinal Analysis " *
trial " OR " random * trial " OR " comparative study
treatment " OR RCT OR " program * evaluation " OR
analysis " OR " project appraisal " OR " pr
" propensity score matching " OR " propensity
design " OR " difference * in " OR " meta analysis " OR " di
" meta analy * " OR " control * random * trial * " OR
allocation * " OR " instrument " OR " assessment * " OR " systematic literature review *
" controlled study " OR " comparative study " " Quasi-experiment * " OR " quasi ex
" experimental group " OR " control community " OR " intervention commun * " OR " control
communities " OR " intervention condition * " OR " control * condition * " OR " control participant * "
OR " experimental condition * " OR counterfactual OR " discontinu * design " OR " fixed effect * " OR
" double differenc * " OR " panel data " OR " double robust " OR " pipeline approach " OR " pipeline
method " OR " pipeline comparison " OR " impact assessment " OR " econometric analys * " OR " cross-
sectional data " OR " fixed effect * " OR " rapid e
" counterfactual " OR " counterfactual " OR " counter-factual " OR " control * evaluation " OR
" randomized field " OR " randomised field "

C. INTERVENTIONS

1. ACTIVE CHOICE, COMMITMENTS AND GOAL SETTING

"choice architecture" OR "active choice" OR "proxy measure" OR "advance direct" OR "checklists" OR "default setting" OR "reference point" OR "framing" OR "commitment"

2. INCENTIVES AND LOTTERIES

"incentive" OR "reward" OR "award" OR "gift" OR "lottery" OR "penalty" OR "reinforcement"

3. PRIMING, FEEDBACK, REMINDERS AND SALIENCE

"priming" OR "nudge" OR "nudging" OR "advice" OR "answer" OR "solution pointer" OR "label" OR "salience" OR "reminder" OR "messenger"

4. SOCIAL NORMS AND BENCHMARKING, RULE OF THUMB

"norm" OR "social proof" OR "herd mentality" OR "goal setting" OR "benchmarking" OR "peer comparison" OR "heuristic" OR "rule of thumb"

5. COGNITIVE BEHAVIOURAL THERAPY AND REDUCING HASSLES

"cognitive behavioural therapy" OR "intelligence" OR "introspective activity" OR "procedural barrier" OR "processual barrier"

6. BEHAVIOURAL SCIENCE CONCEPTS

"behaviour science" OR "behaviour economics" OR "insight" OR "behavior science" OR "behaviour insight" OR "action bias" OR "ambiguity aversion" OR "bounded rationality" OR "certain overload" OR "Chunking" OR "cognitive dissonance" OR "decision fatigue" OR "diversification bias" OR "Hedonic adaptation" OR "Homo economicus"

Appendix 5. DATA EXTRACTION TOOL

DESCRIPTION	QUESTION
Date when form was completed	<i>Date when form was completed</i>
ID of person extracting data	<i>ID of person extracting data</i>
Publication title	<i>Title of publication</i>
Linked studies	<i>Surname of first author</i>
Other papers used for coding	<i>If there is any study linked to this one</i>
Publication date	<i>Year (letter – if more than one study from that author and that year)</i>
Publication type	<i>What is the impact evaluation publication type?</i>
Publication ID	<i>Publication ID</i>
Funding agency	<i>Who is funding the evaluation/study?</i>
Name of funding agency	<i>Please add name of the agency funding the evaluation</i>
Independence of evaluation	<i>What level of independence is there between the implementing agency and study team?</i>
Independent data collection	<i>Has the data been collected by an independent party?</i>
Conflict of interest	<i>Is there a potential conflict of interest associated with study which could influence results collected/reported? (e.g. is there a declaration of conflict of interest? Are any of the authors related in any way to the funding or implementing institution?)</i>
Language of publication	<i>Language of publication of the impact evaluation, e.g. Spanish, English etc.</i>
Other methods	<i>If the impact evaluation addresses questions other than effectiveness, note questions and methods used here.</i>
Country	<i>List countries the study was conducted in</i>
Detailed location	<i>If provided, give detailed information on where the study took place within a country, for example regions/districts covered</i>
World Bank region	Select region(s) in which the study was conducted, according to the World Bank. For more info on region classification see http://data.worldbank.org/country
World Bank income category	<i>Select the World Bank income classification of the country at the time of the study</i>
Programme or project name	<i>State the programme or project name. If no name, then list the location</i>
Intervention type	<i>Select the intervention type</i>
Description of intervention(s)	<i>Describe the selected intervention</i>
Personnel implementing the programme/origin of intervention	<i>Who was in charge of implementing the programme? State origin of intervention:</i> <i>Community-based</i> <i>Non-governmental organization (NGO)</i> <i>Local/national government</i>

DESCRIPTION	QUESTION
	<i>Foreign government</i> <i>Other</i>
Objectives of intervention	<i>State any objectives stated in study or other document</i>
Intervention implementing agency	<i>Who is implementing the intervention? State the name (and department) of the implementing agency.</i>
Implementation funding agency	<i>Type of funder for the implementation of the intervention</i>
Intervention funding agency	<i>Name of intervention funding agency</i>
Intervention target group	<i>Select the target intervention group</i>
Targeting methods	<i>How were beneficiaries targeted for the programme (e.g. how was the targeting implemented)?</i>
Intervention start	<i>Start date (if not stated, state study date) of intervention</i>
Intervention end	<i>State end date (if ongoing, state ongoing)</i>
Intervention length	<i>State intervention length (months)</i>
Consideration of equity	<i>Does the study consider equity?</i>
Equity methods	<i>How does the study consider equity?</i>
Equity dimension	<i>What dimension(s) of equity does the study consider? PROGRESS + indicators (multiple choice – may pick more than one)</i>
Information about programme take-up	<i>Is there any information about programme take-up? Commentary by authors should be used when information on programme take-up, etc., is not backed up by some sort of research, or when the authors do not report that/how they collected data to assess these areas.</i>
Methods of assessing take-up	<i>Which methods are used to assess programme take-up?</i>
Results of the assessment of take-up	<i>What is the result/information provided of the assessment of programme take-up?</i>
Information about programme adherence (among beneficiaries)	<i>Is there any information about programme adherence (among beneficiaries)? Commentary by authors should be used when information on programme adherence, etc., is not backed up by some sort of research, or when the authors do not report that/how they collected data to assess these areas.</i>
Methods of assessing adherence	<i>Which methods are used to assess programme adherence for beneficiaries? This includes attrition and dropout rates, adherence to appointments, etc.</i>
Results of the assessment of adherence	<i>What is the result/information provided of the assessment of programme adherence?</i>
Information about implementation fidelity/intervention delivery quality (among implementers)	<i>Is there any information on implementation fidelity/intervention delivery quality? Commentary by authors should be used when information on programme adherence, etc., is not backed up by some sort of research, or when the authors do not report that/how they collected data to assess these areas.</i>
Methods of assessing intervention fidelity	<i>Which methods are used to assess implementation fidelity/intervention delivery quality by the implementing partner?</i>

DESCRIPTION	QUESTION
Results of the assessment of intervention fidelity	<i>What is the result/information provided of the assessment of implementation fidelity/intervention delivery quality?</i>
Incentives	<i>Were incentives provided to intervention participants?</i>
Other descriptions of process/implementation factors	<i>Any other description of process/implementation factors not covered above</i>
Results	<i>Report here any material relevant to causal mechanisms and barriers and enablers</i>
Cost	<i>Are any unit cost data/cost-effectiveness estimates provided?</i>
Cost details	<i>If yes, report any details of unit cost and/or total cost. Please also report the year and currency</i>
Length of study	<i>Length of study in months (where study length not reported, code as length of intervention, noting that in brackets)</i>
Efficacy or effectiveness trial	<i>Was the intervention implemented in real world conditions? By real world we mean a programme implemented independently of the evaluation, either by government, NGO or international agency</i>
Sampling frame for the study	<i>State the sampling frame (list of all those within a population who can be sampled, i.e. households, communities) for selection of study participants (i.e. census, etc)</i>
Author discussion of external validity	<i>Do the authors discuss or explicitly address generalizability/applicability?</i>
Programme theory	<i>Do the authors make explicit reference to programme theory, theory of change or similar?</i>
	<i>Report any description/statement of programme theory as stated by author(s).</i>
	<i>Is the study using theory to inform the evaluation design and analysis?</i>
Outcome type (outcome category)	<i>Definition of outcome</i>
Outcome timing	<i>1 to 3 years More than 3 years Can't tell</i>
Number of outcomes	<i>State the number of outcomes</i>
Timing of outcome measurement	<i>Only after Before and after Can't tell</i>
Methods	<i>Brief description of the estimation methods</i>
Commentary on methods (if multiple methods are selected)	<i>State here if multiple methods are selected</i>
Study design	<i>Choose the type of study RCT Regression discontinuity Matching/propensity score matching (PSM)</i>

DESCRIPTION	QUESTION
	<i>Instrumental variable/2SLS</i> <i>Difference in difference</i> <i>Interrupted time series</i> <i>Controlled before and after</i> <i>Heckman</i>
Target population gender	<i>State here the gender-targeted population</i> <i>Female</i> <i>Male</i> <i>Female and male</i>
Target population age	<i>Indicate the population</i> <i>Young adults (18-35)</i> <i>Adults (36-65)</i> <i>Elderly (65+)</i> <i>Mixed</i> <i>Not specified</i>
Target population income	<i>Indicate here the target population income</i> <i>Low</i> <i>Middle</i> <i>Diverse</i> <i>Not specified</i>
Target population living environment	<i>State the target population living environment</i> <i>Rural</i> <i>Urban</i>
Target population specific restrictions	<i>Please provide details</i>
Number of treatment arms	<i>Choose a number</i>
Number of intervention components	<i>Choose a number from 1-5</i>
Effect size calculations	
Direction of the effect	<i>Effect favours treatment</i> <i>Effect favours comparison</i> <i>Zero effect</i> <i>C a n ' t t e l l</i>
What intervention (if any) did the comparison group receive?	<i>No treatment</i> <i>As usual</i> <i>Alternative intervention</i> <i>Other</i> <i>C a n ' t t e l l</i>
Were there any differences in the measurement of this outcome between	<i>Yes</i> <i>No</i>

DESCRIPTION	QUESTION
the treatment group participants and the comparison?	
Effect is statistically significant?	Yes No Can't tell
Treatment sample size	<i>Insert treatment sample size here</i>
Control sample size	<i>Insert control sample size here</i>
Nature of the measures	Continuous Dichotomous Hand-calculated data
The following group of questions applies only if "Nature of the	
Treatment group mean	<i>Insert numerical value</i>
Comparison group mean	<i>Insert numerical value</i>
Are means reported above adjusted?	Yes No
Treatment group standard deviation	<i>Insert numerical value</i>
Comparison group standard deviation	<i>Insert numerical value</i>
Treatment group standard error	<i>Insert numerical value</i>
Comparison group standard error	<i>Insert numerical value</i>
T-value from an independent t-test	<i>Insert numerical value</i>
The following group of questions applies only if "Nature of the	
Treatment group – number of participants who experienced a change	<i>Insert numerical value</i>
Comparison group – number of participants who experienced a change	<i>Insert numerical value</i>
Treatment group – proportion of participants who experienced a change	<i>Insert numerical value</i>
Comparison group – proportion of participants who experienced a change	<i>Insert numerical value</i>
Are the proportions above adjusted for pre-test variables?	Yes No
Logged odds-ratio	<i>Insert numerical value</i>
Standard error of logged odds-ratio	<i>Insert numerical value</i>
Logged odds-ratio adjusted?	Yes No

DESCRIPTION	QUESTION
Chi-square with $df=1$ (2 by 2 contingency table)	<i>Insert numerical value</i>
Correlation coefficient	<i>Insert numerical value</i>
The following group of questions applies to all studies:	
Hand calculated d-type effect size	<i>Insert numerical value</i>
Hand calculated error of the d-type effect size	<i>Insert numerical value</i>
Hand calculated odds-ratio effect size	<i>Insert numerical value</i>
Hand calculated odds-ratio standard error	<i>Insert numerical value</i>
Intermediate outcomes or themes (knowledge, skills)	<i>State intermediate outcomes or themes here</i>
Questions applying to all studies:	
Are there results coming from regressions?	<i>Yes</i> <i>No</i>
Sample size	<i>Insert sample size here</i>
The following group of questions applies only if there are results coming from regressions:	
Method: Econometric model?	<i>Insert numerical value</i>
Standard deviation effect	<i>Insert numerical value</i>
Effect (mean)	<i>Insert numerical value</i>
Controls	<i>Insert numerical value</i>
Standard deviation: Y	<i>Insert numerical value</i>
Standard deviation: X	<i>Insert numerical value</i>
B (beta)	<i>Insert numerical value</i>
Standard error B (beta)	<i>Insert numerical value</i>
Degrees of freedom	<i>Insert numerical value</i>
Data type	<i>Panel</i> <i>Cross-section</i> <i>Time-series</i>

Appendix 6. CRITICAL APPRAISAL TOOL

METHODOLOGICAL APPRAISAL CRITERIA				RESPONSE		
				Yes	No	Comment
<p><i>If randomized control trial, start after confounding bias. For all other study designs, start here.</i></p> <p>I. <u>Bias in selection of participants for the study</u> <i>Are participants selected in a way that minimizes selection bias?¹¹</i></p> <p><u>Appraisal indicators</u> Consider whether:</p>						
<p><i>i. There is an adequate description of how and why a sample was chosen (i.e. identified/selected/recruited)</i></p>						
<p><i>ii. There is adequate sample size to allow for representative and/or statistically significant conclusions</i></p>						
<p><i>iii. Participants in the control¹² group were sampled from the same population as that of the treatment</i></p>						
<p><i>iv. Group allocation process minimized potential risk of bias (e.g. using computer algorithms)</i></p>						
<p><i>v. The selection of participants into the study (or into the analysis) is based on participant characteristics observed after the start of the intervention</i></p>						
Low risk of bias	Risk of bias	High risk of bias	Critical risk of bias	<i>Worth it to continue: Y/N?</i>		
<p>II. <u>Bias due to confounding</u> <i>Is confounding potentially controllable in the context of this study?</i></p> <p><u>Appraisal indicators:</u> Consider whether:</p>						
<p><i>i. There is potential for confounding of the effect of the intervention in this study. If yes, provide example of confounding domain in comment box¹³</i></p>						
<p><i>ii. Where matching was applied, it featured sufficient criteria¹⁴</i></p>						

¹¹ Selection bias can occur both in the way that individuals are accepted for participation in a study, and in the way that "treatment" is assigned to individuals once they have been accepted into a study. This section deals with both these understandings of selection bias.

¹² The terms "control" and "comparison" group refer to any group with which the treatment of interest is compared, and that are presumed to represent conditions in the absence of that treatment, whether a true random control or not.

¹³ Confounding domains are those for which, in the context of this study, adjustment is expected to lead to an important change in the estimated effect of the intervention.

¹⁴ Matching can be done on the calculated propensity score or covariates. If the latter, it should ideally be done on pre-test measures and other characteristics, such as demographic. Answer "no" if the study only matched on pre-test measures of some or all variables used later as outcome measures OR matched only on online characteristics.

METHODOLOGICAL APPRAISAL CRITERIA				RESPONSE		
				Yes	No	Comment
iii. Where relevant, the authors conducted an appropriate analysis that controlled for all potential/remaining critical confounding domains after matching had been applied						
iv. The authors avoided adjusting for variables identified after the intervention has been administered						
v. The treatment and control group are comparable after matching/controls have been done. Select one of the following: <input type="checkbox"/> No statistically significant differences <input type="checkbox"/> Statistically significant differences <input type="checkbox"/> Negligible descriptive differences <input type="checkbox"/> Significant descriptive differences						
Low risk of bias	Risk of bias	High risk of bias	Critical risk of bias	Worth it to continue: Y/N?		
<p><i>If randomized control trial, skip I + II (above) and start here.</i> <u>Bias due to confounding (because of ineffective randomization)</u> <i>Is allocation of treatment status truly random?</i> <u>Appraisal indicators</u> Consider whether:</p>						
i. Eligibility criteria for study entry are specified						
ii. There is a clear description of the randomization process and methods are robust						
iii. The unit of randomization and number of participants is clearly stated (pay special attention to treatment and control locations/ balance)						
iv. Characteristics of both baseline and endline sample are provided ¹⁵ and at endline the treatment and control groups are comparable. Select one of the following: <input type="checkbox"/> No statistically significant differences <input type="checkbox"/> Statistically significant differences <input type="checkbox"/> Negligible descriptive differences <input type="checkbox"/> Significant descriptive differences						
Low risk of bias	Risk of bias	High risk of bias	Critical risk of bias	If critical risk of bias, treat as non-random study		
<p>III. <u>Bias due to departures from intended interventions</u> <i>Was the intervention implemented as laid out in the study protocol?</i> <u>Appraisal indicators</u> Consider whether:</p>						

¹⁵ Preferable condition: A RCT with appropriate randomization procedure can be included without showing baseline data, as both experimental groups can be assumed to be equal at baseline by design.

METHODOLOGICAL APPRAISAL CRITERIA				RESPONSE		
				Yes	No	Comment
i. <i>The critical co-interventions were balanced across intervention and control groups</i>						
ii. <i>Treatment switches were low enough not to threaten the validity of the estimated effect of the intervention</i>						
iii. <i>Implementation failure was minor and unlikely to threaten the validity of the estimated effect of the intervention</i>						
iv. <i>It is possible that the intervention was adopted by the controls (contamination and possible crossing-over)¹⁶</i>						
v. <i>It is possible that knowledge of group allocation affects how the two study groups are treated during delivery and evaluation of the intervention¹⁷</i>						
Low risk of bias	Risk of bias	High risk of bias	Critical risk of bias	Worth it to continue: Y/N?		
IV. <u>Bias due to missing/incomplete data (attrition)</u> <i>Are the intervention and control groups free of critical differences in participants with missing/incomplete data?</i> <u>Appraisal indicators</u> Consider whether:						
i. <i>Outcome data are reasonably complete (80% or above)¹⁸</i>						
ii. <i>If level of attrition (or other forms of missing/incomplete data) is more than 20%, are reasons for the missing data reported?</i>						
iii. <i>If level of attrition (or other forms of missing/incomplete data) is more than 20%, do the authors demonstrate similarity between remaining participants and those lost to attrition, and are the proportion of participants with missing/incomplete data and reasons for missing/incomplete data similar across groups?</i>						
iv. <i>If level of attrition (or other forms of missing/incomplete data) is more than 20%, were appropriate statistical methods used to account for missing data? (e.g. sensitivity analysis)¹⁹</i>						
v. <i>If not possible to control for missing/incomplete data, are outcomes with missing/incomplete data excluded from analysis?</i>						
Low risk of bias	Risk of bias	High risk of bias	Critical risk of bias	Worth it to continue: Y/N?		

¹⁶ Whilst challenging in terms of estimating impact, the presence of spill-overs might be an important finding in itself.

¹⁷ Consider only in extreme cases in which preferential treatment is clearly evident; blinding is generally not expected in social interventions.

¹⁸ The assumption here is that the level of attrition (or other forms of missing/incomplete data) is sufficiently low to not require adjustment.

¹⁹ Select "no" if the study addresses missing/incomplete data through simple estimates of missing data and observations.

METHODOLOGICAL APPRAISAL CRITERIA				RESPONSE		
				Yes	No	Comment
V. <u>Bias in measurement of outcomes</u> <i>Are measurements appropriate, e.g. clear origin, or validity known?</i> <u>Appraisal indicators</u> Consider whether:						
i. <i>There was an adequate period for follow up²⁰</i>						
ii. <i>The outcome measure (e.g. employment status, income) was clearly defined and objective²¹</i>						
iii. <i>Outcomes were assessed using standardized instruments and indicators</i>						
iv. <i>Outcome measurements reflect what the experiment set out to measure</i>						
v. <i>The methods of outcome assessment were comparable across groups</i>						
vi. <i>Outcome assessors were aware of the intervention received by study participants?²²</i>						
Low risk of bias	Risk of bias	High risk of bias	Critical risk of bias	Worth it to continue: Y/N?		
VI. <u>Bias in selection of results reported</u> <i>Are the reported outcomes consistent with the proposed outcomes at the protocol stage?</i> <u>Appraisal indicators</u> Consider whether:						
i. <i>It is unlikely that the reported effect estimate has been selected for publication due to it being a particularly notable finding among numerous exploratory analyses</i>						
ii. <i>It is unlikely that the reported effect estimate is prone to selective reporting from among multiple outcome measurements within the outcome domain</i>						
iii. <i>It is unlikely that the reported effect estimate is prone to selective reporting from among multiple analyses of the outcome measurements, including sub-group analysis</i>						
iv. <i>If sub-group/ancillary/adjusted analyses are presented, are these pre-specified or exploratory?</i>						

²⁰ In many social science interventions, follow-up is not required to coincide with the start of the treatment; further, longer periods of follow-up are often required to measure changes.

²¹ Subjective measures (e.g. those based on self-report) are likely to have lower reliability and validity than objective measures.

²² Consider only in extreme cases in which preferential treatment is clearly evident; blinding is generally not expected in social interventions.

METHODOLOGICAL APPRAISAL CRITERIA				RESPONSE		
				Yes	No	Comment
v. <i>The analysis includes an intention to treat analysis. (If so, was this appropriate and were appropriate methods used to account for missing data?)²³</i>						
vi. <i>Do the authors report on all variables they aimed to study (as specified in their protocol or study aims/research questions)?</i>						
Low risk of bias	Risk of bias	High risk of bias	Critical risk of bias			
Overall risk of bias:						

²³ Usually in clinical RCTs, rare in social science: only rate if conducted.

Appendix 7. GRADING OF RECOMMENDATIONS ASSESSMENT, DEVELOPMENT AND EVALUATION (GRADE) TOOL (EXAMPLE)

CERTAINTY ASSESSMENT						SAMPLE SIZE	EFFECT	CERTAINTY	IMPORTANCE
Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations		Absolute (95% CI)		
Outcome 1									
RCT	Serious	Serious	Not serious	Serious	None	737	SMD 0.02 SD higher (0.12 lower to 0.16 higher)	VERY LOW	Limited importance
Outcome 2									
RCT - 3	Serious	Serious	Not serious	Not serious	None	4,991	SMD 0.14 SD higher (0.01 higher to 0.28 higher)	LOW	Important, but not critical
Outcome 3									
RCT - 6 QED - 2	Very serious	Not serious	Not serious	Not serious	None	9,970	SMD 0.09 SD higher (0.02 higher to 0.16 higher)	LOW	Important, but not critical
Outcome 4									
RCT	Very serious	Serious	Not serious	Not serious	None	3,219	Two negative and three positive effect estimates with a 95% CI range of -0.08 to 0.16	VERY LOW	Important, but not critical
Outcome 5									
RCT	Very serious	Serious	Not serious	Not serious	None	3,219	SMD 0.02 SD higher (0.09 lower to 0.05 higher)	VERY LOW	Important, but not critical
Outcome 6									

CERTAINTY ASSESSMENT						SAMPLE SIZE	EFFECT	CERTAINTY	IMPORTANCE
Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations		Absolute (95% CI)		
RCT	Not serious	Serious	Not serious	Serious	None	3,543	Five positive effect estimates with a 95% CI range of -0.00 to 0.41	LOW	Important, but not critical
Outcome 7									
RCT - 7	Serious	Serious	Not serious	Not serious	None	8,359	SMD 0.06 SD higher (0.02 lower to 0.14 higher)	LOW	Critical
Outcome 8									
RCT - 2 QED - 1	Very serious	Serious	Not serious	Not serious	None	5,233	SMD 0.14 SD higher (0.02 higher to 0.26 higher)	VERY LOW	Limited importance

Independent Evaluation Unit
Green Climate Fund
175, Art center-daero, Yeonsu-gu
Incheon 22004, Republic of Korea
Tel. (+82) 032-458-6450
ieu@gcfund.org
<https://ieu.greenclimate.fund>

