# Systematic review of conservation interventions to promote voluntary behavior change

Laura Thomas-Walters[1]*, Jamie McCallum[2], Ryan Montgomery[3], Claire Petros[4], Anita K. Y. Wan[5], & Diogo Veríssimo[4,6]

1. Oregon State University; lathowal@gmail.com

2. Force for Nature, UK; jamie.mccallum@forcefornature.org

3.ryan.montgomery@arguk.org

4. University of Oxford, UK;

5. Sun Yat-Sen University, China

6. On The Edge Conservation, UK

*Corresponding author: email lathowal@gmail.com

Keywords: behavior change, human behavior, nudging, OHAT, social marketing, social science

Article impact statement: There is strong evidence that education, prompts, and feedback interventions result in proenvironmental behavior change.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/cobi.14000.

## Abstract

Humanity is having ever greater impacts on the environment. Understanding human behaviour is vital to developing interventions that effectively lead to pro-environmental behaviour change, whether the focus is at the individual or societal level. However, interventions in many fields have historically lacked any robust form of evaluation, which makes it hard to be confident that these conservation intervention have successfully helped protect the environment. We conducted a systematic review to assess how effective non-pecuniary and non-regulatory intervention have been in changing environmental behaviour, adapting the Office of Health Assessment and Translation methodology. We started with more than 300,000 records and after critical appraisal of quality identified 128 individual studies that merited inclusion in the review. We classified interventions by thematic area, type of intervention, the number of times audiences were exposed to interventions, and the length of time for which interventions ran. Most studies reported a positive effect (n=96), while the next most common outcome was no effect (n=28). Few studies reported negative (n=1) or mixed (n=3) effects. We found strong evidence that education, prompts and feedback interventions can result in positive behaviour change. The strongest evidence comes from combining multiple interventions in one programme. Neither exposure duration nor frequency had an effect on the likelihood of desired behaviour change. We found comparatively few studies that tested the effects of voluntary interventions on non-Western populations, or that measured actual conservation behaviours. While prompts and education are well-studied, we are lacking evidence to support the use of conservation devices and demonstrations. There is a clear need to both improve the quality of impact evaluation conducted, and the reporting standards for intervention results.

## Introduction

Humanity is having ever greater impacts on the environment, and those impacts are driven by human decision making (Lewis & Maslin, 2015). Many people lead unsustainable lifestyles, particularly in higher-income countries, contributing to major environmental issues like climate change and biodiversity loss (Cowling, 2014; Fischer et al., 2012). Problematic behaviours, such as excessive water and energy use, need to be addressed urgently (OECD, 2011). Accordingly, conservation as a discipline has increasingly embraced the social sciences to aid in the design and evaluation of behaviour change interventions (Bennett et al., 2017; Moon et al., 2019). Understanding human behaviour is vital to developing interventions that mitigate threats to the environment and effectively lead to pro-environmental behaviour change, whether the focus is at the individual or societal level. However, these interventions have historically lacked robust evaluation, which makes it hard to know whether (and how) conservation intervention have helped protect the natural world (Curzon & Kontoleon, 2016; Josefsson et al., 2020; Junker et al., 2020).

Traditional responses to the environmental crisis have been mainly policy-based (Lucas et al., 2008; Science and Technology Select Committee, 2011). Central among them have been legislation that eliminates or restricts choice and fiscal incentives or disincentives (Lucas et al., 2008; Taylor et al., 2013). For example, governments have implemented restrictions on the disposal of waste, and charges for single-use carrier bags (Goodstein & Polasky, 2014; Poortinga et al., 2013). While important, these policies can be resource-intensive and require political will to implement (Allcott, 2011; Schubert, 2017). They may be politically unpopular as they are intrusive and involve the loss of liberty (although the

restriction of environmental harm may benefit the liberty of people in society more widely; Science and Technology Select Committee, 2011). Pecuniary interventions also require consistent funding in the long-term to be sustainable, and raises questions around autonomy and power, especially in socioeconomically disadvantaged groups (Marteau et al., 2009). Non-coercive approaches to behaviour change have received increasing interest because people retain the freedom to make the choice they wish without concern for legal or financial repercussions, and reliance on political will is lessened (Greenfield & Veríssimo, 2018; Schubert, 2017).

Research testing the effectiveness of these approaches have been conducted for decades (e.g., Asch & Shore, 1975; Krauss et al., 1978), but there still is not a cohesive body of evidence to guide policymakers and practitioners. This could be due to publication in multiple disciplines, including but not limited to social marketing, environmental education, and behavioural economics (Veríssimo & Wan, 2018; Hungerford & Volk, 1990; Lehner et al., 2016). Syntheses of voluntary interventions in the environmental field do exist, but they tend to be narrative reviews, include only specific evidence types like randomised control trials (RCTs), focus on a select thematic area such as energy consumption, or test a specific intervention type like education (e.g., Abrahamse et al., 2005; Abrahamse & Steg, 2013, Byerly et al., 2018; Heimlich & Ardoin, 2008; Nisa et al., 2019; Schubert, 2017; Wolske et al., 2020; Wynes et al., 2018). Reviews also often include proxies for behaviour change, such as changes in behavioural intentions or attitudes. While these proxies have a role in research, their correlation with behaviour is not strong. For example, meta-analyses show that intentions account for only 28% percent of the variance in prospective measures of

behaviour (Sheeran, 2002). This systematic review focuses on actual behaviours with clear environmental impacts and incorporates a broader range of experimental designs and intervention types, as well as having a broader scope that explicitly includes the grey literature. It also includes a rigorous quality assessment process to ensure only robust methodologies are part of the final synthesis.

Systematic reviews synthesise a body of evidence to explore specific research questions. They are the most reliable and comprehensive statement about what works, providing useful information for policymakers and practitioners (Johnson & Hennessy, 2019; Munn et al., 2018). The transparency and rigour of systematic reviews can be enhanced by following a set of accepted principles such as the Cochrane and Campbell Collaboration (2013) standards or the Office of Health Assessment and Translation (OHAT; Rooney et al., 2014) framework.

The OHAT approach is a systematic review methodology to increase transparency, consistency and efficiency in summarizing environmental health-based findings, with the additional goal of improving data management and display (OHAT 2014). It draws on the best public health protocols (e.g., PRISMA, PECOTS, Campbell Collaboration) while being able to cope with the broader set of conditions and wide range of data types required in the wider environmental health sciences. For example, it allows for the inclusion of relevant and high-quality papers in the grey literature, to help minimise publication bias (Savoie et al., 2003). It also embraces experimental designs beyond RCTs, an important factor for

environmental reviews. It is not always feasible or appropriate to perform an RCT, and in some areas (such as biodiversity conservation) there are very few we can learn from. Including only these designs therefore excludes a large body of evidence (Christie, 2020). Moreover, there are a variety of alternative, rigorous, quasi-experimental designs, using techniques such as matching, synthetic control, or regression discontinuity to control for observed and unobserved covariates, that are comparable in levels of bias to RCTs (Christie et al., 2019; Pynegar et al., 2019). RCTs are also vulnerable to biases, for example linked to randomisation failure or differential attrition (Jadad & Enkin, 2007). Unless there are other risk factors (e.g., see Step 5 in the Methods), rigorous quasi-experimental designs can be treated with a similar level of confidence to RCTs.

In this systematic review, we explore how effective non-pecuniary and non-regulatory intervention have been in changing environmental behaviours. To do so, we include only studies measuring actual behaviour. In particular, we focus on the quality and rigour of the evidence base. We examine the strength of evidence that different types of interventions, such as feedback or goal-setting, will result in desired behaviour change. We also identify important gaps in the literature.

## Methods

We adapted the OHAT seven-step framework for systematic reviews, as documented below (Akers et al., 2009; Rooney et al., 2014; Fig 1). The breadth and inclusive nature of the process enabled us to create some degree of standardisation across studies that varied in

experimental design and outcome measurement. We sent the review protocol that we developed to two external experts for feedback before the start of the study. As a result of this feedback, minor changes to the protocol were made, including the addition of keywords and clarification of the scope of the review.

### Step 1 – Problem formulation and protocol development

We wanted to cast a wide net to include all non-regulatory and non-pecuniary interventions which tried to solve an environmental threat by changing human behaviour. Studies may have been published in fields as distinct as waste management, sustainable transport, or social marketing. We specifically focussed on assessing the quality of the evidence base and identifying where there were gaps. We were also interested in direct behavioural measures, and not just proxies such as attitudes, intentions, or self-reported behaviour. Outcomes were behaviours with a clear environmental impact, such as water or energy consumption, travel mode choice (e.g., public transport versus car journeys), recycling participation (frequency and volume), or littering.

### Step 2 – Literature search, processing and selection

Between February and May 2015 we systematically searched across multiple databases to identify high quality and relevant studies from both the peer-reviewed and grey literature. We developed our keyword search strategy based on electronic searches of bibliographic databases or platforms, project funding databases and specialised internet search engines and repositories (summary in Table 1, full details in SI 1). The selected search terms

provided broad coverage of environmental behaviours and approaches, while keeping the focus on actual behaviours. No language or date restrictions were applied to the searches, although only English keywords were used. ISI Web of Science and EBSCOhost are platforms which provide access to a wide range of bibliographic databases, and the specific databases we accessed are listed in full in the SI 2. Other electronic databases and indexes for peer-reviewed literature included:

- SciVerse's Scopus
- International Bibliography of the Social Sciences
- PsycINFO
- Google Scholar
- Education Resources Information Center
- Environmental Evidence
- Campbell Collaboration systematic review database
- International Initiative for Impact Evaluation review database

For grey literature we also searched:

- ProQuest Digital Dissertations and Theses
- PolicyFile
- My Environmental Education Resource Assistant

- Canadian Evaluation Society Unpublished Literature Bank
- System for Information on Grey Literature in Europe
- CORDIS Library
- Fostering Sustainable Behaviour: Community Based Social Marketing
- Tools of change
- Rufford Foundation project database
- Conservation Leadership Award project database
- Rainforest Alliance Eco-Index
- Darwin Initiative Project Database

Boolean operators (i.e., “AND” and “OR”) were used as appropriate. For each database, the number of hits per search phrase in titles and abstracts were recorded. The number of records retrieved for the largest bibliographic databases/platforms are listed in SI 3. There were additional searches for both British and American English, i.e., ‘behaviour’ and ‘behaviour’. All searches were noted and tracked on a Microsoft Excel spreadsheet for reference.

All records, including the full texts, were screened manually by two co-authors, AKYW and DV, who split the workload in half. Duplicates in the initial database of records were automatically identified through EndNote and then confirmed manually. We triaged the

studies in three stages: title, abstract and full-text review, and only included empirical, primary data studies. At each stage, studies were categorised as 'Accept', 'Maybe' or 'Reject'. This was based on the explicit inclusion/exclusion criteria outlined in Table 2. Two reviewers independently tested the reliability of the triage process using 100 random records. We calculated agreement using Cohen's kappa coefficient, with a cut-off criterion for substantial agreement above 0.6 (McHugh, 2012).

### Step 3 – Extract data from studies

Once we had a complete dataset of eligible studies, we extracted data for each record (SI 4). All studies were listed with administrative identifiers including record source, title, first author and publication year. We categorized them into one of six themes from the Community-Based Social Marketing classification (https://cbsm.com/; McKenzie-Mohr & Schultz, 2014) – agriculture and conservation, energy, transportation, water, waste and pollution, or a mix of multiple themes. As well as PECOTS (population, exposure, comparison, outcomes & time), further information on the intervention context, design and measure of outcomes were retrieved for each article to obtain data for an overall understanding of evaluation measures undertaken by project organizers.

At this stage, we also developed a taxonomy of different intervention types. We drew on existing behavioural intervention taxonomies (namely Abraham & Michie, 2008; Dolan et al., 2012; Kok et al., 2016; Michie et al., 2015; Michie et al., 2013), but tailored it to the types of interventions that were present in the review studies. There were six main types: education, demonstrations, conservation devices, feedback, goal-setting, and prompts (Table 3).

### Step 4 – Initial quality assessment of individual studies

Following OHAT guidelines we ranked each study based on i) quality of reporting, ii) relevance of experimental design to outcome, and iii) risk of bias (Rooney et al., 2014). Quality of reporting refers to how well a study was completed or reported. Relevance refers to the relevance of experimental design to the behavioural outcome. Risk of bias includes external validity or directness and applicability (i.e., how well a study addresses the topic under review). The first two criteria were graded from 1 (low) to 3 (high), while risk of bias was graded from 1 (definitely high) to 4 (definitely low). We then added the three grades for a total score out of 10. At this point, any study which scored less than six was removed from the review as it would be considered problematic in multiple key aspects of study quality (Office of Health Assessment and Translation, 2015).

### Step 5: Confidence rating for studies

At this stage in the OHAT protocol, similar studies would be clumped together to enable the processing of a large number of papers and to determine common threads. In our review, there are multiple ways in which we could categorise the studies (for example, thematic and intervention types). As a relatively small number of papers made it to this stage (Fig 2) and there is considerable heterogeneity in experimental design and behavioural outcomes, we rated each study individually.

We developed a confidence rating for each study based on the presence or absence of key features identified in the OHAT process (Schünemann et al., 2011). Studies earned one point for each of the following features that would increase our confidence in the study result:

i. *Exposure to intervention controlled by researchers*. The ability to largely eliminate confounding by randomising the allocation of exposure.

ii. *Exposure prior to outcome*. Exposure to intervention occurred prior to outcome measurements.

iii. *Individual outcome data*. Outcome measurements were collected at the individual level.

iv. *Comparison group used*. A comparison or control group was used within the study.

v. *Large magnitude of effect*. Studies that achieve >50% magnitude of effect relative to the control group within the population of a study received a +1.

vi. *Plausible intervention-outcome relationship*. If there is a degree of plausibility between the level of exposure and outcome, then it is more likely that the result did not occur due to chance. Is the degree of change in a population subject to the degree of a given exposure? This is especially relevant when looking at studies that vary degrees of the same exposure or in factorial study design, when multiple exposures are being applied in different combinations.

vii. *Residual confounding*. This refers to effect modification that would bias the effect estimate towards the null. Conversely, studies reporting no effect and remaining consistent across studies would move the effect estimate from the null. A score of +1

was awarded if the replicates were robust and comparable but with extremely variable results, while 0 was given if the treatment, intervention, and control were deemed to have low applicability.

viii. *Consistency amongst control and treatment populations*. This refers to extreme similarity in a population, notably the robustness of the replicates, comparable controls, intervention and treatment.

However, the presence of any of these features that would decrease our confidence in the study result meant a one-point deduction:

i. *Risk of bias*. This was extrapolated from the OHAT Step 4 risk of bias score, where anything that scored 'probably high' or 'definitely high' qualified for the point deduction.

ii. *Unexplained inconsistency*. This referred to the external validity or indirect measures of the behavioural outcome, obtained by reading the results and discussion.

iii. *Indirectness.* This was assessed using the relevance scores from the OHAT Step 4, where low relevance qualified for the point deduction.

iv. *Imprecision*. This is the degree of certainty surrounding an effect estimate, and was assessed based on sample size, the power of the statistical methods used and their confidence intervals. For example, large standard deviations (i.e., the standard deviation is greater than the mean) or an odds ratio where the ratio of the upper to lower 95% confidence intervals is greater than 10.

As we were still rating individual studies rather than a body of evidence, we excluded the fifth feature suggested by OHAT, publication bias. We summed up the points for a maximum score of 8 and a minimum score of -4. Following OHAT protocol any study which scored 0 or less was removed from the review at this point, as we would have "very low confidence" in their outcomes (Office of Health Assessment and Translation, 2015). We were then able to give the remaining studies a confidence rating using these scores, "high" (>5), "moderate" (3-4) or "low" (1-2) confidence. 30 studies were independently reviewed by a second rater, and Cohen's kappa coefficient showed substantial agreement (0.67; McHugh, 2012).

### Step 6: Translate confidence ratings into evidence of desired behaviour change

We extracted data on the behavioural outcomes of each study, noting whether they resulted in positive behaviour change or negative/no change. We then classified the level of evidence for desired behaviour change that each study provided according to their confidence ratings and direction of effect (Table 4). This strategy involved the use of three terms to describe the level of evidence for behavioural outcomes, "known", "presumed", and "suspected", which were directly translated from the confidence-in-the-evidence ratings. As there were only four studies with negative or mixed results and they did not significantly impact our findings, we focussed on evidence that a given variable would lead to positive or *desired* behaviour change.

We were then able to collate the results from multiple studies to calculate the overall level of evidence for a given variable by calculating the mean of the numerical rating from different

studies (from 3 to -3). If the mean confidence rating for a group of studies is less than 1, behaviour change is suspected. If it is between $1 \geq X > 2$, behaviour change is presumed. If it is between $2 \geq X > 3$, behaviour change is known. These terms were taken from the OHAT protocol (Office of Health Assessment and Translation, 2015). We also sought to calculate an effect size for different intervention types. However, the number of papers that reported these or enough information to allow for their calculation was insufficient to allow for a meaningful analysis. As such we did not include effect sizes in this analysis.

### Data analysis

The OHAT protocol acknowledges that disparate exposure and outcome assessments may preclude formal statistical meta-analysis and therefore does not specify statistical tests for outcomes (Office of Health Assessment and Translation, 2015). However, we conducted exploratory data analysis to further understand the relationships between variables. Our aim was to examine the effects of different variables on a) the distribution in the number of studies, b) variation in the quality of studies, and c) the strength of evidence that a given intervention results in desired behaviour change. Quality of studies refers to the confidence rating for each individual study, while strength of evidence is based on the mean scores calculated for a body of studies as described in step 6. We were interested in whether there were differences in these outcomes across time, space and thematic area. The type of intervention, the number of times audiences were exposed to the intervention, and the length of time for which the intervention ran were also potential variables of interest.

Data analysis was conducted in R version 4.0.0, and we selected the appropriate statistical tests based on the dependent and independent variables. We employed a Kendall Tau-b correlation to test the distribution in the number of studies across time, and Chi-Squared tests were used to examine the distribution in the number of studies across all the other variables of interest (region, theme, intervention type, duration of intervention, and the number of exposures to interventions). We calculated a Spearman's rank correlation coefficient to test whether there was a variation in the quality of studies across time, while we used a Kruskal-Wallis test for distribution in quality of studies across the remaining variables of interest. We calculated a Spearman's rank correlation coefficient to test whether the strength of evidence that a given intervention will result in desired behaviour change varies across time, and again we used a Kruskal-Wallis test for the other variables.

## Results

From an initial identification of 338,408 records, we found 128 individual studies published in 107 articles that met our quality criteria for inclusion (Fig 2). Confidence ratings for these studies varied from low (n=27), moderate (n=55), and high (n=48). Most studies reported a positive effect (n=96), while the next most common outcome was no effect (28). Few studies reported negative (n=1) or mixed (n=3) effects. Unfortunately, only 25 (19%) of the 128 interventions reported enough detail in the statistical results to calculate a standardised effect size for meta-analyses.

### Date of publication

Studies in this review had been published between 1975 to 2015. The variation in the number of studies published over this period is not statistically significant ($z = 1.36$, $p = 0.17$).

There has however been a significant improvement in the quality of studies conducted by publication year ($r_s = 0.27$, $p = 0.002$), and the strength of evidence that a given intervention would result in desired behaviour change has also increased with time ($r_s = 0.21$, $p = 0.02$).

### Location of intervention

There is an uneven distribution of studies across continents ($\chi^2(5) = 172.19$, $p < 0.001$), with a disproportionate number located in North America and Europe. There is also a significant variation in the quality of studies across regions ($\chi^2(5) = 15.99$, $p = 0.007$), with Europe featuring an above-average proportion of high-quality studies. The level of evidence that a given intervention will result in desired behaviour change also varies by study location ($\chi^2(5) = 22.62$, $p < 0.001$). Those conducted in Asia, Europe, and Oceania are more likely to find strong evidence that an intervention will result in the desired behaviour change (Fig 3).

### Study theme

Studies are unequally distributed across study themes ($\chi^2(5) = 55.75$, $p < 0.001$), and a disproportionate number focus on waste and energy: agriculture and conservation (5), energy (40), transport (17), waste (41), water (15), or a mixture of these (10). We are unable to reject the null hypothesis that there is no relationship between the quality of the studies and study theme ($\chi^2(5) = 3.57$, $p = 0.61$), however, nor can we reject the null hypothesis that there is no relationship between the likelihood of a given intervention resulting in behaviour change and the study theme ($\chi^2(5) = 9.78$, $p = 0.08$).

371 Intervention type

372 We found an uneven distribution of studies across intervention types ($\chi2(6) = 65.92$, $p <$
373 0.001), with a disproportionate number focussing on education interventions and very few
374 looking at demonstrations: conservation device (8), demonstration (3), education (45),
375 feedback (24), goal setting (10), prompt (24), or a combination of the different types (14).
376 The combinations are:

377 - 4 - Feedback, goal setting
378 - 3 - Education, conservation device
379 - 3 - Feedback, education
380 - 1 - Feedback, goal setting, and coaching
381 - 1 - Education, demonstration
382 - 1 - Prompt, conservation device
383 - 1 - Education, prompt, demonstration

384

385 We are unable to reject the null hypothesis that there is no relationship between the quality
386 of the studies in this review and the intervention type ($\chi2(6) = 7.74$, $p = 0.26$), nor can we
387 reject the null hypothesis that there is no relationship between the likelihood of a given
388 intervention resulting in behaviour change and the intervention type ($\chi2(6) = 6.86$, $p = 0.33$).
389 However, we can assess the body of evidence for each type individually. Based on the
390 quality and reported outcomes of studies examining interventions based on education,

prompts and feedback, and their outcomes, we can presume that desirable behaviour change will result from this intervention type (Fig 4). For goal-setting desirable behaviour change is only suspected. We did not rate interventions based on conservation devices and demonstrations due to the low number of studies focusing on these intervention types. However, our highest confidence rests on the use of multiple different intervention types, for which positive change is a known outcome.

### Intervention duration and exposure frequency

We identified interventions that ran from 10 minutes to multiple years, and audiences who may have been exposed to the intervention once, multiple times, or continually (e.g., a conservation device installed in the home). There was a significant variation in the number of studies by both intervention length ($\chi2(5) = 37.52$, $p < 0.001$) and nature of exposure ($\chi2(2) = 26.82$, $p < 0.001$), with a disproportionate amount lasting either less than one day or between one and three months, and involving multiple exposures. We did not find any evidence that study quality varies by intervention duration ($\chi2(5) = 9.2$, $p = 0.1$), or the number of times an audience was exposed to an intervention ($\chi2(2) = 0.86$, $p = 0.65$). Overall, we cannot reject the null hypotheses that there is no relationship between the likelihood of a given intervention resulting in behaviour change and the quantity of exposures ($\chi2(2) = 1.78$, $p = 0.41$; Fig 5) or the duration of the intervention ($\chi2(5) = 8.18$, $p = 0.15$).

The number of studies published varies significantly across all variables of interest except for publication date (Table 5; full details in text). However, the only variables for which we are able to detect a relationship with study quality and evidence for desired behaviour change are date of publication and location.

### Ineligible studies

Finally, we also explored trends in the studies that were removed for lacking controls or being of too low quality. We found that the publication of studies including controls has decreased over time ($r_s = -0.19$, $p = 0.03$). However, in the subset that do include a control, quality has improved ($r_s = 0.23$ $p = 0.01$). We are unable to reject the null hypothesis that there is no relationship between presence of a control ($\chi^2(1) = 0.31$, $p = 0.58$) or quality ($\chi^2(5) = 5.51$, $p = 0.36$) and grey versus peer-reviewed literature.

## Discussion

We identified strong evidence that education, prompts and feedback interventions can result in positive behaviour change. Given the current evidence base, we can conclude that combining multiple interventions in one campaign is most likely to lead to the desired outcomes. However, we still need to know what combinations of intervention types work best, for what behaviours, in what contexts, and for what duration and intensity. Surprisingly, we were unable to reject the null hypothesis that there is no relationship between the overall duration or frequency of exposure to the intervention and the likelihood of effecting positive behaviour change. However, the quality of the studies and the strength of the existing

evidence overall vary by both date of publication and location of study. This may be due to greater resources available to researchers in the West (Waldron et al., 2013).

Geographic location was the only variable in our review to significantly impact the distribution, quality, and evidence strength of studies. This continental bias towards North America and Europe has unfortunately also been identified in other reviews of conservation research, and limits our ability to draw firm conclusions about interventions' effectiveness beyond these locations (Wilson et al., 2016). Previous research in behavioural science shows that there is substantial variability in experimental results across populations, and a lack of cultural diversity in research participants could skew responses to interventions (Henrich, Heine and Norenzayan, 2010).

Results from the systematic review show an average of three high-quality studies published each year globally. At the same time, nearly two-thirds of eligible studies (223) had to be removed from further review as they failed to include an independent control. A rigorous impact evaluation should provide credible evidence by using an appropriate counterfactual to establish causal attribution (Ferraro, 2009). Control groups act as a counterfactual, thereby mitigating bias in the comparison of impacts when bias in allocation has been taken into account. Less than 20% of included studies presented enough statistical information to determine effect size. Unfortunately, the need to incorporate better-designed measurement protocols into environmental behaviour change interventions is another conclusion that repeatedly arises in reviews of the literature (Byerly et al., 2018; Delmas et al., 2013).

Indeed, Byerly et al. (2018) recently noted that many studies are poorly designed, lacking adequate controls and sufficient statistical power. Considering the exponential spread of anthropogenic threats to the environment and the urgency for effective mitigation strategies it is vital that we improve the rigour with which we approach impact evaluation, and pay this issue the attention it deserves (Cowling, 2014; Fischer et al., 2012). This widespread and systemic failure of the field should be a call to action for all conservation social scientists.

*Narrative synthesis*

We conducted a narrative synthesis to explore major themes and relationships between and within these studies, in order to identify any factors contributing to their reported success or failure (Popay et al., 2006). The cohort of studies span a forty-year period, and more recent studies build upon findings from earlier benchmark studies. For example, Carrico's intervention design (Carricio, 2005) focusing on motivational feedback was informed by Becker's findings that goal commitment without feedback is ineffective (Becker, 1978). Further, most of the studies were informed by the large body of current behavioural theories in the social sciences, such as nudge theory (Baca-Motes, 2013, Baird, 2014), moral norms (Ayres, 2012; Thørgesen, 1997), the theory of planned behaviour (Thørgesen, 2009), habit hypothesis (Bamberg 2006) and motivational feedback (Becker 1978).

The design of an intervention seemed strongly related to the problem it was intended to address. For example, researchers working on energy consumption tended to employ multi-faceted interventions including the provision of educational material (Carrico 2009), individual consumption feedback (Hayes 1977), and peer feedback and goal setting (Loock

2013), whereas those focused on litter prevention and waste management tested more visual interventions such as information via signs and prompts (Sussman 2013, Hansmann 2003).

The exposure of the targeted audience to the interventions varied widely, ranging from a one-off exposure event to months or even years, as well as variability in the number of times the intervention was implemented. There appears to be little consensus on what are appropriate exposure times for eliciting a behavioural response. Several studies noted the importance of continuous follow-up or lack of follow-up and lengthening exposure time to increase likelihood of a successful outcome (Baca-Motes 2013; Baird 2014; Harrigan 1994; Harrigan 1994; Hayes 1977), yet few studies included long-term monitoring of behaviours. When studies did conduct long-term monitoring, the initial reported positive behavioural changes were often found to diminish over time, perhaps suggesting a need for consistency in interventions over a longer period of time in order to achieve the desired long-term behavioural change. This is despite the likely attrition of participants who have ceased behaviour change, compared to those who maintained the desired behaviour change. Unfortunately, the variability between studies, such as the length of the intervention to the follow-up duration limits our ability to suggest any guidelines on a best practice timeframe for future research.

Rarely, outcomes besides the intended pro-environmental behaviour were noted. Unintended detrimental effects following interventions were cited in several studies. For example, increased energy usage after receiving peer feedback (Ayers 2012). Different

reasons were put forward to explain these undesirable outcomes, including the 'boomerang effect' and 'moral licensing' (Ayres 2012, Nomura 2011). Tiefenbeck (2013) recorded moral licensing when participants in a water conservation campaign reduced their water consumption as intended, but also increased their electricity use.

Unintended or additional outcomes can also be beneficial, such as in the case of positive spillover (Evans 2012; Haq 2008). For example, environmental messages that promote car-sharing for reasons other than personal benefit, may also lead to an uptake in recycling (Evans 2012). However, cross-domain adoption of additional pro-environmental behaviours as a byproduct of interventions was not commonly measured (n=6).

Finally, one common theme that emerged throughout this review is that impactful or novel studies tended to include strategies that made personal connections between the broader issues and the targeted audience. For example, eliciting emotional responses; building empathy or aligning with an individual's internal standards (Hansmann 2003). Strategies that lead to an emotional reaction can result in a positive behavioural change (Sussman 2013). These approaches result in individual-led action, linked to high-motivation methods and making cognitive connections. This suggests that if an intervention is thought-provoking and connects with audiences on a higher cognitive level, it is more likely to result in positive behaviour change (Hansmann 2003; Miller 2009).

521 Comparison with other reviews

522 During the initial literature search we tried to be as inclusive as possible. We used broad
523 search terms, and searched in multiple bibliographic databases and languages (Haddaway
524 & Macura, 2018). We also included a wide breadth of journals from different disciplines,
525 although a more formal effort to benchmark search comprehensiveness using a set of
526 papers already identified within a category would have allowed us to better understand the
527 degree to which we are capturing all the relevant literature. What distinguishes our review
528 from previous work is the broad screening criteria for experimental design and subject focus,
529 and the thorough critical appraisal process we used to classify all the studies that met
530 inclusion criteria. Instead of vote-counting based on statistical significance or the exclusive
531 on one kind of experimental design such as randomised controlled trials, we weighted the
532 studies by quality to measure the strength of evidence (Haddaway & Macura, 2018). In
533 addition, any study which did not meet a certain threshold for rigour was removed from the
534 final analysis. This means that all the studies in our review still meet key standards of
535 robustness, which is important as previous reviews have found behavioural effect sizes
536 varying with study rigour (Delmas et al., 2013). We also included a wide range of grey
537 literature, which is often overlooked in other reviews due to concerns about study quality.
538 However, we found no significant differences in quality between the grey and peer-reviewed
539 literature.

540

541 Previous reviews have also suggested that although single policy tools frequently fail to
542 reduce household energy consumption, synergistic effects can come from combining
543 interventions (Dietz et al., 2009). For example, the most effective interventions for daily

energy-use behaviours generally involve a mixture of mass-media messages, household-specific information, and social influences. Indeed, combining feedback with goal setting has been shown to work particularly well in the energy sector (Abrahamse et al., 2005). This echoes a notable finding from our review, that we have strong evidence to support the effectiveness of multiple interventions, such as feedback + goal setting, and education + conservation devices.

Previous reviews of behaviour change interventions in fields such as public health have suggested that intervention success may be linked to intensity. The weight of evidence shows campaigns with longer duration or more frequent contact time can lead to positive outcomes such as greater weight loss or a reduction in sexual risk behaviours (Chandra-Mouli et al., 2015; Greenhow, 2011; Robin et al., 2004). However, interventions often vary considerably in duration and delivery, preventing even a descriptive analysis let alone the identification of an optimal formula (Durlak & DuPre, 2008; McCoy et al., 2010; Wei et al., 2011). We found no clear link between intervention success and duration of exposure, but there may be heterogeneity amongst different intervention types. This should be a priority area for future research.

### Methodological considerations

As with all syntheses, there is the possibility that the studies we identified were subject to the "file-drawer effect", or a bias towards the publication of studies with positive/significant results (Franco et al., 2014; Scargle, 2000). For example, publication bias has been shown

in previous reviews of behavioural science literature (Francis, 2012). A visual assessment of outcome distributions (the number of studies published with positive rather than negative or no change) certainly suggests that publication bias may be a concern in Asia, Africa, Europe, and Oceania, but less so in North and South America (Fig 3). Further, the absence of any published studies showing negative results for intervention types such as conservation devices or demonstrations is alarming, and limits what we can say about the true effectiveness of these interventions.

Research from the medical field shows that when studies are pre-registered, negative outcomes are more likely (Dwan et al., 2008). Moves by journals such as Conservation Biology to allow pre-registration are a step in the right direction to address publication bias. For this review, we tried to mitigate publication bias by searching both the peer-reviewed and grey literature (Haddaway & Macura, 2018). We also included a wide breadth of journals from different disciplines. Our focus on direct behavioural outcomes likely restricted eligible studies to topics where behaviour can feasibly be measured. This does not mean that interventions have been ineffective in changing more elusive behaviours such as consumption of illegal wildlife trade products, but rather we do not yet have enough evidence to come to a confident conclusion (Veríssimo & Wan, 2018). In addition, grouping diverse behaviours, from water use to transport choices, may have masked interesting trends in the relative effectiveness of different interventions (Heimlich & Ardoin, 2008). Currently, however, the lack of studies prevents a more detailed analysis of these possible interactions.

Literature searches were conducted in 2015, and resources have not been available to update the initial search. This suggests an important trade-off between the comprehensiveness of a search strategy and the ability to produce syntheses in a timely manner. Future reviews should consider carefully the extent of their scope, using tools such as forward and backward citation search as a potential way to ensure targeted searches still remain comprehensive.

Since 2015, multiple high-quality, rigorous studies have been published that would have met the criteria for our review (Schwartz et al., 2020; Weigel et al., 2021; Wolstenholme et al., 2020). It is possible that including these studies would have improved the average robustness of the research featured in this review. However, it is also worth noting that the latest literature on the use of behavioural science to conserve biodiversity continues to identify most, if not all, of the challenges we highlighted above, including lack of controls, narrow geographic focus, and failure to measure actual behaviours (Balmford et al., 2021; Nilsson et al., 2020; Palm-Forster et al., 2019).

## Conclusion

Several key gaps in the literature need to be addressed. We found comparatively few studies that tested the effects of voluntary interventions on non-Western populations, or that measured actual conservation behaviours. While prompts and education are well-studied, we are lacking evidence to support the use of conservation devices and demonstrations. Future researchers should aim to fill these gaps, and should also improve reporting

standards. More detail is needed both on the statistical front to enable the calculation of effect sizes and in terms of intervention implementation. For example, we had to use a coarse categorisation scheme for the duration and frequency of exposure analyses. If we had the quantity of information for more fine-grained analysis, our results would be more robust. Finally, we need to investigate the extent to which behaviour change persists after the intervention has ceased (Burns & Savan, 2018; Byerly et al., 2018)

We demonstrate there is strong evidence that a range of different, well-designed intervention types can result in desired behaviour change. Notably, the strongest evidence comes from the combination of multiple intervention types, for example, both conservation devices and education. Encouragingly, we found successful interventions across a range of durations and exposures, indicated that behaviour change can occur from even short-term efforts. This is not to neglect the role of governments and industry in addressing major environmental issues, but instead highlights some of the effective approaches they can use to maximise impact. The findings from this review should be used by practitioners to guide future interventions, and by researchers to inform future studies.

*Acknowledgments*: This work was funded by the David H. Smith Conservation Research Fellowship. We are grateful to have received three thoughtful and in-depth reviews on an early draft of the manuscript.

References

Abraham, C., & Michie, S. (2008). A taxonomy of behavior change techniques used in interventions. Health Psychology, 27(3), 379–387.

Abrahamse, W., & Steg, L. (2013). Social influence approaches to encourage resource conservation: A meta-analysis. *Global Environmental Change*, *23*(6), 1773–1785.

Abrahamse, W., Steg, L., Vlek, C., & Rothengatter, T. (2005). A review of intervention studies aimed at household energy conservation. Journal of Environmental Psychology, 25(3), 273–291.

Akers, J., Aguiar-Ibáñez, R., Baba-Akbari Sari, A., et al. (2009). CRD's guidance for undertaking reviews in health care.

Allcott, H. (2011). Social norms and energy conservation. Journal of Public Economics, 95(9–10), 1082–1095.

Allcott, H., & Rogers, T. (2014). The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation. American Economic Journal, 104(10), 3003–3037.

Asch, J., & Shore, B. M. (1975). Conservation behavior as the outcome of environmental education. Journal of Environmental Education, 6(4), 25–33.

Balmford, A., Bradbury, R. B., Bauer, J. M., et al. (2021). Making more effective use of human behavioural science in conservation interventions. Biological Conservation, 261(July), 109256. https://doi.org/10.1016/j.biocon.2021.109256

652 Bennett, N. J., Roth, R., Klain, S. C., et al. (2017). Mainstreaming the social sciences in
653 conservation. Conservation Biology, 13(1), 56–66.

654 Brown, T. J., Ham, S. H., & Hughes, M. (2010). Picking up litter: An application of theory-
655 based communication to influence tourist behaviour in protected areas. Journal of
656 Sustainable Tourism, 18(7), 879–900.

657 Burns, S., & Savan, B. (2018). The Post-Intervention Persistence of Energy Conservation
658 Behaviors: An Evaluation of the 'Start Green' Program. Sustainability, 10(3), 809.

659 Byerly, H., Balmford, A., Ferraro, P. J., et al. (2018). Nudging pro-environmental behavior:
660 evidence and opportunities. Frontiers in Ecology and the Environment, 16(3), 159–
661 168.

662 Chandra-Mouli, V., Lane, C., & Wong, S. (2015). What Does Not Work in Adolescent Sexual
663 and Reproductive Health: A Review of Evidence on Interventions Commonly
664 Accepted as Best Practices. Global Health: Science and Practice, 3(3), 333–340.

665 Christie, A. P. (2020). Quantifying and addressing the prevalence and bias of study designs
666 in the environmental and social sciences. Nature Communications, 11(6377).

667 Christie, A. P., Amano, T., Martin, P. A., et al. (2019). Simple study designs in ecology
668 produce inaccurate estimates of biodiversity responses. Journal of Applied Ecology,
669 56(12), 2742–2754.

670 Cotterill, S., John, P., Liu, H., & Nomura, H. (2009). Mobilizing citizen effort to enhance
671 environmental outcomes: A randomized controlled trial of a door-to-door recycling
672 campaign. Journal of Environmental Management, 91(2), 403–410.

673 Cowling, R. M. (2014). Let's get serious about human behavior and conservation.
674 Conservation Letters, 7(3), 147–148. https://doi.org/10.1111/conl.12106

675 Curzon, H. F., & Kontoleon, A. (2016). From ignorance to evidence? The use of programme
676 evaluation in conservation: Evidence from a Delphi survey of conservation experts.
677 Journal of Environmental Management, 180, 466–475.

678 Delmas, M. A., Fischlein, M., & Asensio, O. (2013). Information strategies and energy
679 conservation behavior: A meta-analysis of experimental studies from 1975-2011.
680 Energy Policy, 61, 729–739.

681 Dietz, T., Gardner, G. T., Gilligan, J., Stern, P. C., & Vandenbergh, M. P. (2009). Household
682 actions can provide a behavioral wedge to rapidly reduce US carbon emissions.
683 Proceedings of the National Academy of Sciences, 106(44), 18452–18456.

684 Dolan, P., Hallsworth, M., Halpern, D., King, D., Metcalfe, R., & Vlaev, I. (2012). Influencing
685 behaviour: The mindspace way. Journal of Economic Psychology, 33, 264–277.

686 Dwan, K., Altman, D. G., Arnaiz, J. A., Bloom, J., Chan, A. W., Cronin, E., Decullier, E.,
687 Easterbrook, P. J., Von Elm, E., Gamble, C., Ghersi, D., Ioannidis, J. P. A., Simes, J.,
688 & Williamson, P. R. (2008). Systematic review of the empirical evidence of study
689 publication bias and outcome reporting bias. PLoS ONE, 3(8).

690 Ferraro, P. J. (2009). Counterfactual thinking and impact evaluation in environmental policy.
691 New Directions for Evaluation, 122, 75–84.

692 Fischer, J., Dyball, R., Fazey, I., Gross, C., Dovers, S., Ehrlich, P. R., Brulle, R. J.,
693 Christensen, C., & Borden, R. J. (2012). Human behavior and sustainability. Frontiers
694 in Ecology and the Environment, 10(3), 153–160.

695 Francis, G. (2012). Publication bias and the failure of replication in experimental psychology.
696 Psychonomic Bulletin and Review, 19(6), 975–991.

697 Franco, A., Malhotra, N., & Simonovits, G. (2014). Publication bias in the social sciences:
698 Unlocking the file drawer. Science, 345(6203), 1502–1505.

699 Geller, E. S., Erickson, J. B., & Buttram, B. A. (1983). Attempts to promote residential water
700 conservation with educational, behavioral and engineering Strategies. Population and
701 Environment, 6(2), 96–112.

702 Goodstein, E. S., & Polasky, S. (2014). Economics and the Environment (7th ed.). John
703 Wiley & Sons.

704 Greenfield, S., & Veríssimo, D. (2018). To what extent is social marketing used in demand
705 reduction campaigns for illegal wildlife products? Insights from elephant ivory and
706 rhino horn. Social Marketing Quarterly, 25(1), 40–54.

707 Greenhow, E. H. (2011). Systematic review of reviews of intervention components
708 associated with increased effectiveness in dietary and physical activity interventions.
709 BMC Public Health, 89(2277), 479–481.

710 Haddaway, N. R., & Macura, B. (2018). The role of reporting standards in producing robust
711 literature reviews. Nature Climate Change, 8(6), 444–447.

712 Heimlich, J. E., & Ardoin, N. M. (2008). Understanding behavior to understand behavior
713 change: A literature review. Environmental Education Research, 14(3), 215–237.

714 Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world?
715 Behavioural and Brain Sciences, 33, 61–135.

716 Hungerford, H. R., & Volk, T. L. (1990). Changing learner behavior through environmental
717 education. In Journal of Environmental Education (Vol. 21, Issue 3, pp. 8–21).

718 Jadad A, Enkin M. (2007) Bias in randomized controlled trials. In: Randomized Controlled
719 Trials: Questions, Answers, and Musings. 2nd ed. p. 29–47.

720 Johnson, B. T., & Hennessy, E. A. (2019). Systematic reviews and meta-analyses in the
721 health sciences: Best practice methods for research syntheses. Social Science &
722 Medicine, 233(May), 237–251.

723 Josefsson, J., Hiron, M., Arlt, D., et al. (2020). Improving scientific rigour in conservation
724 evaluations and a plea deal for transparency on potential biases. Conservation
725 Letters, e12726.

726 Junker, J., Petrovan, S. O., Arroyo-rodríguez, V., et al. (2020). A severe lack of evidence
727 limits effective conservation of the world's primates. BioScience, 1–10.

728 Kok, G., Gottlieb, N. H., Peters, G.-J. Y., et al. (2016). A taxonomy of behavior change
729 methods: An intervention mapping approach. Health Psychology Review, 10(3), 297–
730 312.

731 Krauss, R. M., Freedman, J. L., & Whitcup, M. (1978). Field and laboratory studies of
732 littering. Journal of Experimental Social Psychology, 14, 109–122.

733 Lehner, M., Mont, O., & Heiskanen, E. (2016). Nudging – A promising tool for sustainable
734 consumption behaviour? Journal of Cleaner Production, 134, 166–177.

735 Lewis, S. L., & Maslin, M. A. (2015). Defining the Anthropocene. Nature, 519(7542), 171–
736 180.

737 Loock, C., Staake, T., & Thiesse, F. (2013). Motivating energy-efficient behavior with green
738 IS: An investigation of goal setting and the role of defaults. MIS Quarterly, 37(4),
739 1313–1332.

740 Lucas, K., Brooks, M., Darnton, A., & Jones, J. E. (2008). Promoting pro-environmental
741 behaviour: Existing evidence and policy implications. Environmental Science and
742 Policy, 11(5), 456–466.

743 Marteau, T. M., Ashcroft, R. E., & Oliver, A. (2009). Using financial incentives to achieve
744 healthy behaviour. BMJ, 338(7701), 983–985.

745 McHugh, M. (2012) Interrater Reliability: The Kappa Statistic. Biochemia Medica, 22, 276–
746 82.

747 McKenzie-Mohr, D., & Schultz, P. W. (2014). Choosing Effective Behavior Change Tools.
748 Social Marketing Quarterly, 20(1), 35–46.

749 Michie, S, Ce, W., Johnston, M., Abraham, C., Ji, F., & Hardeman, W. (2015). Behaviour
750 change techniques: The development and evaluation of a taxonomic method for
751 reporting and describing behaviour change interventions (a suite of five studies
752 involving consensus methods, randomised controlled trials and analysis of qualitative
753 data. Health Technology Assessment, 19(99).

Michie, S, Richardson, M., Johnston, M., Abraham, C., Francis, J., Hardeman, W., Eccles, M. P., Cane, J., & Wood, C. E. (2013). The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: Building an international consensus for the reporting of behavior change interventions. Annals of Behavioral Medicine, 46(1), 81–95. https://doi.org/10.1007/s12160-013-9486-6

Moon, K., Blackman, D. A., Adams, V. M., et al. (2019). Expanding the role of social science in conservation through an engagement with philosophy, methodology, and methods. Methods in Ecology and Evolution, 10(3), 294–302.

Munn, Z., Stern, C., Aromataris, E., Lockwood, C., & Jordan, Z. (2018). What kind of systematic review should I conduct? A proposed typology and guidance for systematic reviewers in the medical and health sciences. BMC Medical Research Methodology, 18(5), 1–9.

Nilsson, D., Fielding, K., & Dean, A. J. (2020). Achieving conservation impact by shifting focus from human attitudes to behaviors. Conservation Biology, 34(1), 1–10. https://doi.org/10.1111/cobi.13363

Nisa, CF., Bélanger, JJ., Schumpe, BM., & Faller, DG. (2019). Meta-analysis of randomised controlled trials testing behavioural interventions to promote household action on climate change. Nature Communications, 10(1).

OECD. (2011). Greening Household Behaviour: The Role of Public Policy. OECD Publishing.

774 Office of Health Assessment and Translation. (2015). Handbook for conducting a literature-
775 based health assessment using OHAT approach for systemic review and evidence
776 integration. https://ntp.niehs.nih.gov/ntp/ohat/pubs/handbookjan2015_508.pdf

777 Palm-Forster, LH., Ferraro, PJ., Janusch, N., Vossler, CA., & Messer, KD. (2019).
778 Behavioral and Experimental Agri-Environmental Research: Methodological
779 Challenges, Literature Gaps, and Recommendations. Environmental and Resource
780 Economics, 73(3), 719–742.

781 Poortinga, W., Whitmarsh, L., & Suffolk, C. (2013). The introduction of a single-use carrier
782 bag charge in Wales: Attitude change and behavioural spillover effects. Journal of
783 Environmental Psychology, 36, 240–247.

784 Popay, J., Roberts, H., Sowden, A., et al. (2006). Guidance on the Conduct of Narrative
785 Synthesis in Systematic Reviews (ESRC Methods Programme).

786 Pynegar, EL., Gibbons, JM., Asquith, NM., & Jones, JPG. (2019). What role should
787 randomized control trials play in providing the evidence base for conservation? Oryx.

788 Robin, L., Dittus, P., Whitaker, D., et al. (2004). Behavioral interventions to reduce incidence
789 of HIV, STD, and pregnancy among adolescents: A decade in review. Journal of
790 Adolescent Health, 34(1), 3–26.

791 Rooney, AA., Boyles, AL., Wolfe, MS., Bucher, JR., & Thayer, KA. (2014). Systematic review
792 and evidence integration for literature-based environmental health science
793 assessments. Environmental Health Perspectives, 122(7), 711–718.

794 Savoie, I., Helmer, D., Green, CJ., & Kazanjian, A. (2003). Beyond MEDLINE: Reducing bias
795 through extended systematic review search. International Journal of Technology
796 Assessment in Health Care, 19(1), 168–178.

797 Scargle, JD. (2000). Publication bias: The “file-drawer” problem in scientific inference.
798 Journal of Scientific Exploration, 14(1), 91–106.

799 Schubert, C. (2017). Green nudges: Do they work? Are they ethical? Ecological Economics,
800 132, 329–342.

801 Schünemann, H., Hill, S., Guyatt, G., Akl, EA., & Ahmed, F. (2011). The GRADE approach
802 and Bradford Hill’s criteria for causation. Journal of Epidemiology and Community
803 Health, 65(5), 392–395.

804 Schwartz, D., Loewenstein, G., & Agüero-Gaete, L. (2020). Encouraging pro-environmental
805 behaviour through green identity labelling. Nature Sustainability.
806 https://doi.org/10.1038/s41893-020-0543-4

807 Science and Technology Select Committee. (2011). Behaviour Change: Vol. HL Paper 1.

808 Sheeran, P. (2002). Intention-behavior relations: A conceptual and empirical review.
809 European Review of Social Psychology, 12(1), 1–36.

810 Smith, RJ., Salazar, G., Starinchak, J., Thomas-Walters, L., & Veríssimo, D. (2020). Social
811 marketing and conservation. In W. J. Sutherland, P. N. M. Brotherton, Z. G. Davies,
812 N. Ockendon, N. Pettorelli, & J. A. Vickery (Eds.), Conservation Research, Policy and
813 Practice (pp. 309–321). Cambridge University Press.

814 Sussman, R., & Gifford, R. (2013). Be the change you want to see: Modeling food
815 composting in public places. Environment and Behavior, 45(3), 323–343.

816 Taylor, CM., Pollard, SJT., Angus, AJ., & Rocks, SA. (2013). Better by design: Rethinking
817 interventions for better environmental regulation. Science of the Total Environment,
818 447, 488–499.

819 Veríssimo, D., & Wan, AKY. (2018). Characterizing efforts to reduce consumer demand for
820 wildlife products. Conservation Biology, 33(3), 623–633.

821 Waldron, A., Mooers, A., Miller, D., Nibbelink, N., et al. (2013). Targeting global conservation
822 funding to limit immediate biodiversity declines. Proceedings of the National
823 Academy of Sciences, 110(29), 12144–12148.

824 Weigel, C., Harden, S., Masuda, YJ., Ranjan, P., Wardropper, CB., Ferraro, PJ., Prokopy,
825 L., & Reddy, S. (2021). Using a randomized controlled trial to develop conservation
826 strategies on rented farmlands. Conservation Letters, e12803.
827 https://doi.org/10.1111/conl.12803

828 Wilson, K. A., Auerbach, N. A., Sam, K., et al. (2016). Conservation research is not
829 happening where it is most needed. PLoS Biology, 14(3), 1–5.

830 Wolske, K. S., Gillingham, K. T., & Schultz, P. W. (2020). Peer influence on household
831 energy behaviours. Nature Energy, 5(3), 202–212.

832 Wolstenholme, E., Poortinga, W., & Whitmarsh, L. (2020). Two Birds, One Stone: The
833 Effectiveness of Health and Environmental Messages to Reduce Meat Consumption

and Encourage Pro-environmental Behavioral Spillover. Frontiers in Psychology, 11. https://doi.org/10.3389/fpsyg.2020.577111

Wright, A. J., Veríssimo, D., Pilfold, K., et al. (2015). Competitive outreach in the 21st century: Why we need conservation marketing. Ocean & Coastal Management, 115, 41–48.

Wynes, S., Nicholas, K. A., Zhao, J., & Donner, S. D. (2018). Measuring what works: Quantifying greenhouse gas emission reductions of behavioural interventions to reduce driving, meat consumption, and household energy use. Environmental Research Letters, 13(113002).

Yin, T. (2013). The role of a social marketing campaign in promoting environmental behavior change in Yuhe provincial nature reserve, Gansu province, China. MSc Thesis. State University of New York

**Table 1.** Keywords and terms used during the comprehensive literature search.

| |
|---|
| "education" AND OR "conservation" OR "outdoor" OR "ecology" OR "adventure" OR "global" OR "field studies" |
| "social marketing" AND "conservation" OR "biodiversity" OR "species" OR "habitat" OR "wildlife" OR "nature" OR "environment" |

| |
|---|
| “community based conservation” OR “community-based conservation” |
| “integrated conservation and development” |
| “community conservation” |
| “community based natural resource management” OR “community-based natural resource management” |
| “energy conservation” AND “behaviour change” |
| “water conservation” AND “behaviour change” |
| “recycling” AND “behaviour change” |
| “littering” AND “behaviour change” |
| “source reduction” AND “waste” AND “behaviour change” |
| “reducing consumption” AND “behaviour change” |
| “composting” AND “behaviour change” |
| “carpooling” AND “behaviour change” |

| |
|---|
| “fuel efficient vehicles” AND “behaviour change” |
| “walking” AND “behaviour change” |
| “mass transit” AND “behaviour change” |
| “biking” AND “behaviour change” |
| “volunteering” AND “behaviour” |

849

850

851 **Table 2**. Inclusion and exclusion criteria used to determine study eligibility

| | **Inclusion Criteria** |
|---|---|
| Population | No age or geographic restrictions |
| Exposure | The intervention must not be pecuniary or regulatory |
| Comparators | Must include a control |
| | The control must be independent |
| | There should be a rationale detailing why the control is comparable to treatment |
| Outcomes | Must have a behavioural outcome (i.e., not just knowledge, attitudes or |

| | |
|---|---|
| | norms) |
| | The behavioural outcome must be relevant to the environment |

852

853

854

855 **Table 3**. Taxonomy of intervention types.

| Intervention type | Definition | Example |
|---|---|---|
| Education | Involves imparting information to increase knowledge or understanding of a behaviour or issue | Visiting households to discuss the benefits of recycling and the local recycling service (Cotterill et al., 2009) |
| Demonstrations | Model a desired behaviour, enabling audiences to learn by observation | Conspicuously disposing of food waste into the appropriate receptacle in a restaurant (Sussman & Gifford, 2013) |
| Conservation devices | Facilitate the performance of a desired behaviour with new technologies or improved | Supplying more energy-efficient stoves to reduce fuelwood consumption (Yin, 2013) |

| | | |
|---|---|---|
| | services | |
| Feedback | Provides data on personal behaviour, possibly with comparison to a stated goal or the behaviour of others | Home energy reports featuring personalized energy use feedback (Allcott & Rogers, 2014) |
| Goal-setting | Encourages audiences to commit to an explicit behavioural target | Asking energy consumers to set a specific energy-saving goal (Loock et al., 2013) |
| Prompts | Uses environmental or social cues to remind audiences to perform a behaviour | Displaying signs with persuasive messages to remind tourists to pick up litter (Brown et al., 2010). |

856
857
858
859 **Table 4**. Evidence of desired behaviour change

| Confidence rating | Behavioural outcomes | |
|---|---|---|
| | Positive | Negative/none |
| High | Known (3) | Known (-3) |
| Moderate | Presumed (2) | Presumed (-2) |
| Low | Suspected (1) | Suspected (-1) |

**Table 5**. Summary of variable effects on outcomes of interest

| Variable | Number | Quality | Evidence |
|---|---|---|---|
| Publication date | Z = 1.36, p=0.17 | **rs = 0.27, p<0.01** | **rs = 0.21, p=0.02** |
| Location | **χ2(5) = 172.19, p<0.01** | **χ2(5) = 15.99, p<0.01** | **χ2(5) = 22.62, p<0.01** |
| Thematic area | **χ2(5) = 55.75, p<0.01** | χ2(5) = 3.57, p=0.61 | **χ2(5) = 9.78, p=0.08** |
| Intervention type | **χ2(6) = 65.92, p<0.01** | χ2(6) = 7.74, p=0.26 | χ2(6) = 6.86, p=0.33 |
| Intervention length | **χ2(5) = 37.52, p<0.01** | χ2(5) = 9.2, p=0.1 | χ2(5) = 8.18, p=0.15 |
| Number of exposures | **χ2(2) = 26.82, p<0.01** | χ2(2) = 0.86, p=0.65 | χ2(2) = 1.78, p=0.41 |

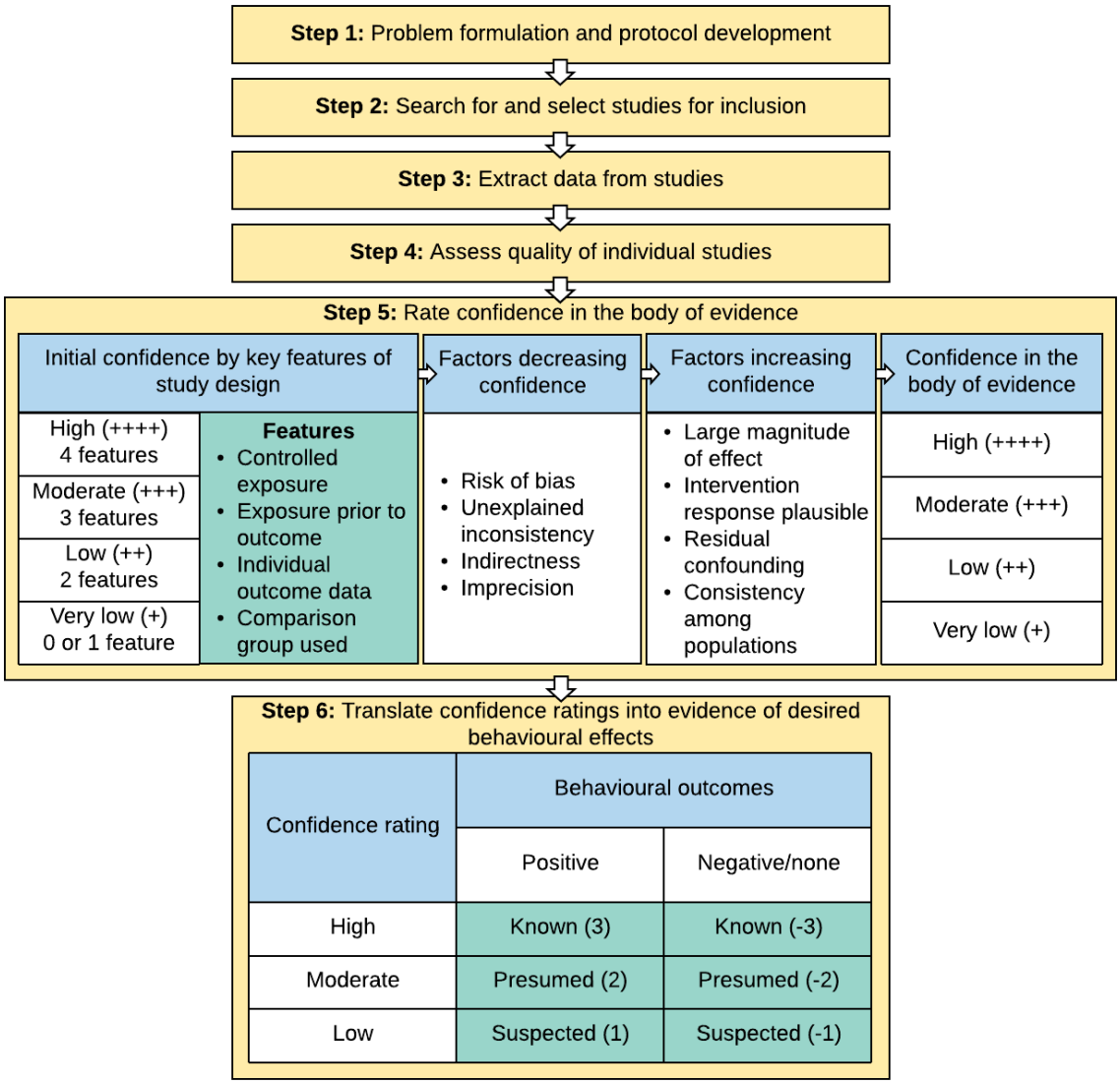



**Figure 1.** Adaptation of OHAT (2014) systematic literature review protocol.

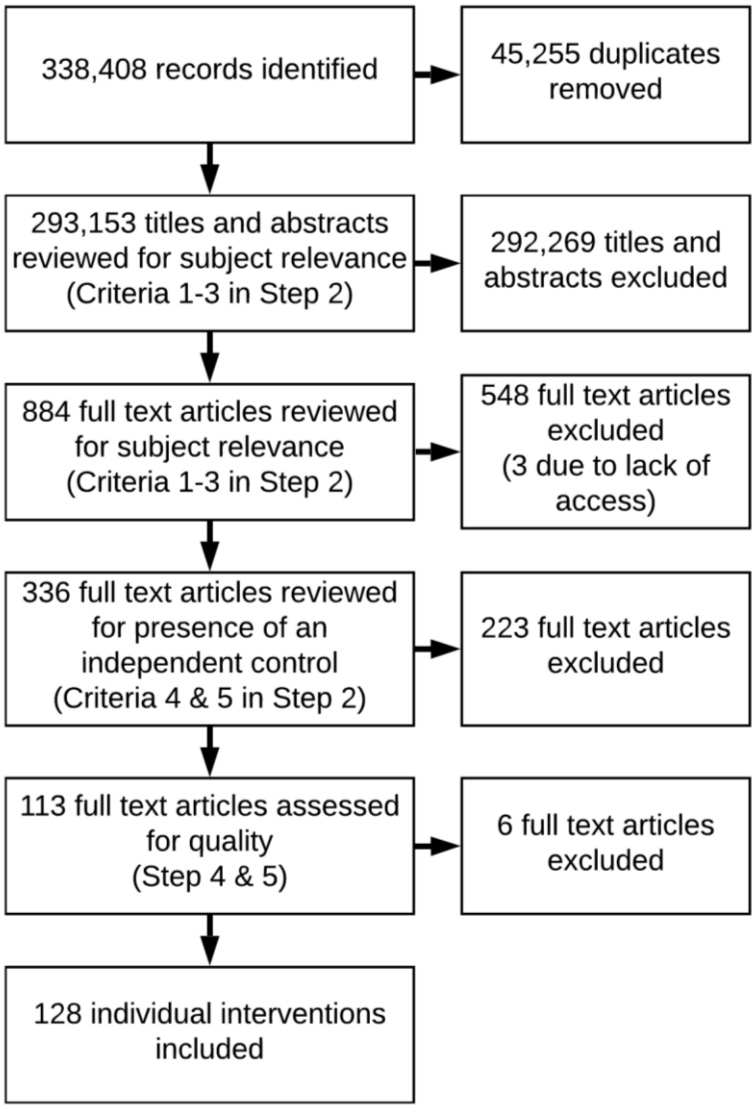
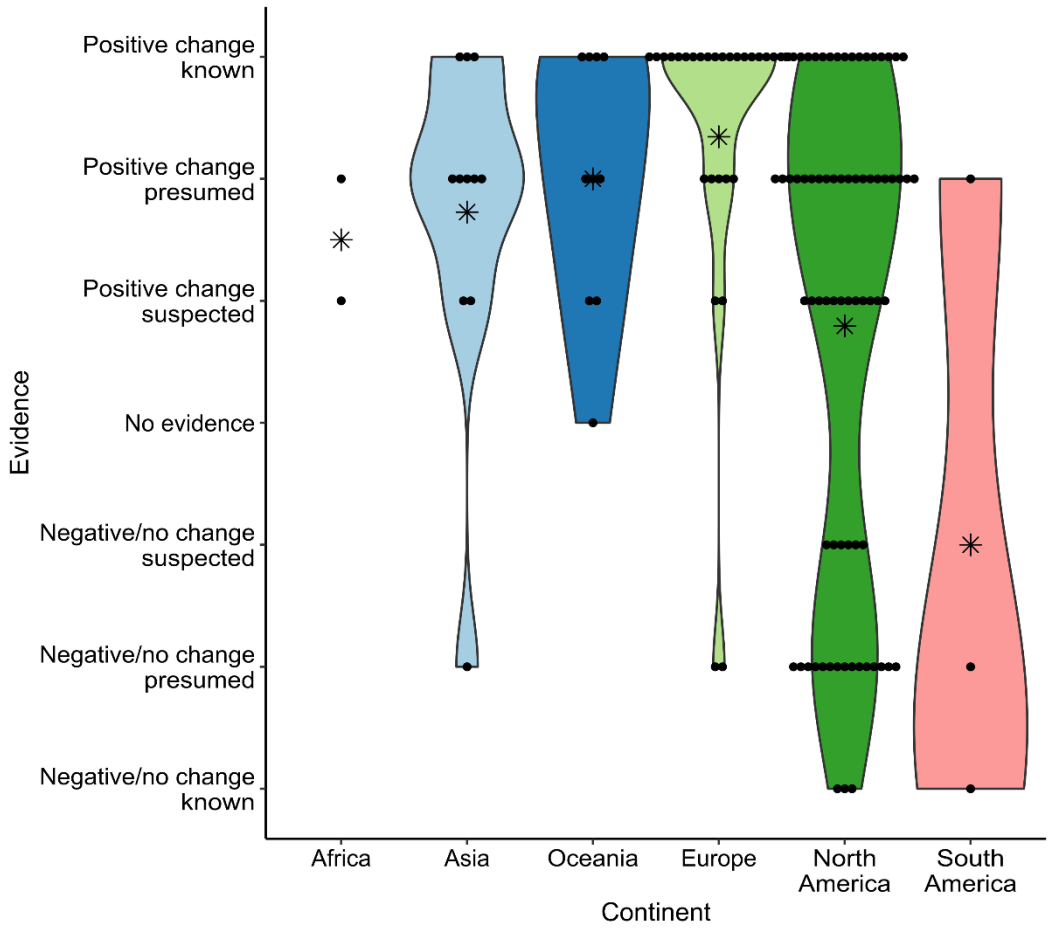



**Figure 2.** Selection of articles for systematic review



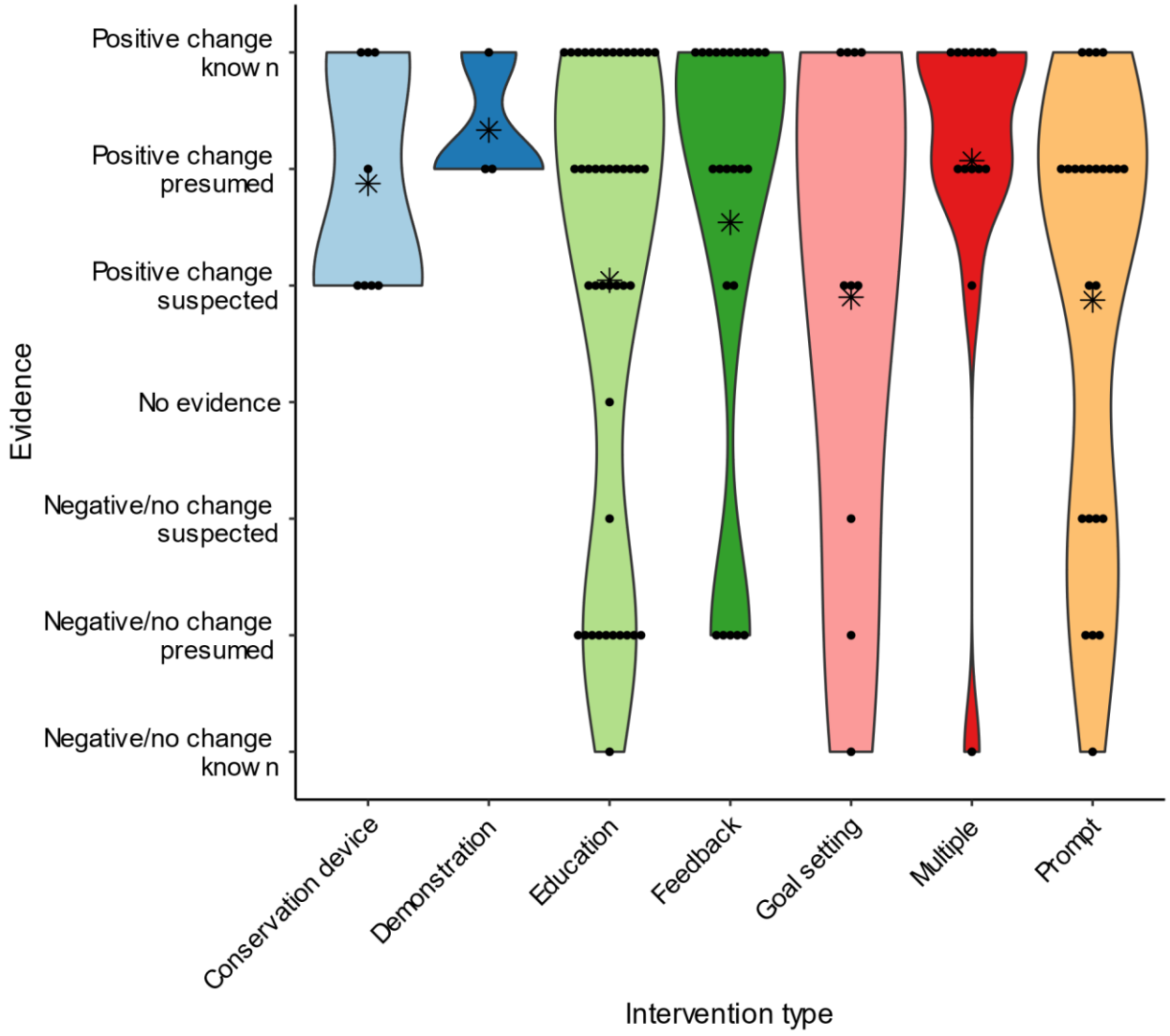

874

875 **Figure 3**. Likelihood that a given intervention will effect desired behaviour change by study location

876

877

878




879

880 **Figure 4**. Likelihood that a given intervention type will result in desired behaviour change, *

881 = overall behavioural categorisation

882

**Figure 5**. Likelihood that a given exposure type or duration length will result in desired behaviour change.