

## Full Length Article

# Payment for ecosystem services and crowding of conservation behavior: A meta-analysis of lab-in-the-field experiments<sup>☆</sup>

Tobias Vorlauffer<sup>a,\*</sup>, Ivo Steimanis<sup>b</sup>, Jan Plassenberg<sup>c</sup>

<sup>a</sup> Leibniz Centre for Agricultural Landscape Research (ZALF), Working Group Governance of Ecosystem Services, Germany

<sup>b</sup> University of Marburg, Working Group Sustainable Use of Natural Resources, School of Business and Economics, Germany

<sup>c</sup> GWS, Institute of Economic Structures Research, Germany

## ARTICLE INFO

## Keywords:

Payments for environmental services  
Conservation  
Motivations  
Behavioral response  
Crowding effects  
Crowding-in

## ABSTRACT

Concerns have been raised that payments for ecosystem services (PES) may crowd out land users' non-monetary motivations to engage in conservation behavior. Especially once incentives are terminated, PES risk to be ineffective or even counterproductive. So far, research has produced mixed evidence. We present the first meta-analysis of studies that investigated crowding effects of PES through lab-in-the-field experiments with 2,894 real-world resource users taking 44,540 conservation decisions. On average, PES are successful in increasing conservation behavior and do not crowd out conservation behavior once incentives have been terminated. Although PES demonstrate greater effectiveness in settings where local resource users directly benefit compared to settings where third parties benefit, there is no evidence suggesting systematic differences in crowding effects between these two situations once PES have been terminated. Based on the available experimental evidence, the frequently voiced risk that PES crowd-out conservation, especially once payments are terminated, cannot be substantiated. However, methodological concerns regarding the internal and external validity of current experiments raise questions about the broader applicability of these findings. This paper outlines potential avenues for future research to address these challenges.

## 1. Introduction

Motivational crowding describes the phenomenon that external monetary incentives can undermine non-monetary motivations to engage in the encouraged behavior. While such crowding effects have been observed and discussed for a wide array of human behaviors (Frey and Jegen 2001; Gneezy, Meier, and Rey-Biel 2011), they have emerged as a key criticism of Payment for Ecosystem Services (PES) (Rode, Gómez-Baggethun, and Krause 2015; Akers and Yasué 2019; Ezzine-de-Blas, Corbera, and Lapeyre 2019; Chan et al. 2017) and resonated in particular among critics of market-based approaches, who see PES as an instrument in line with the increasing commodification of nature (e.g. Kosoy and Corbera 2010; Neuteleers and Engelen 2015; Gómez-Baggethun and Muradian 2015). PES can be defined as economic incentives to land users conditional on providing ecosystem services, for example water regulation or carbon sequestration, or pre-defined land-use practices that are expected to maintain or increase targeted ecosystem services (Wunder 2015). PES have been implemented worldwide in

various contexts (Salzman et al. 2018), and play a pivotal role in agri-environmental policies in Europe and the US (Schomers and Matzdorf 2013). In addition, PES have been considered a key instrument of REDD + for tropical forest conservation and climate change mitigation (Angelsen et al. 2017).

If PES lead to a crowding-out of non-monetary motivations, PES would be less effective than the pure monetary incentive effect suggests. However, most PES programs do not intend or cannot indefinitely provide payments, for example due to budgetary, or political constraints. As a result, the cumulative PES impact could be negative over time when pro-environmental behavior targeted by PES falls below levels that would have been observed without PES (Ezzine-de-Blas, Corbera, and Lapeyre 2019). So far, the debate whether and under which conditions PES lead to a crowding-out of non-monetary motivations has not been settled. While some studies find evidence in favor of the crowding-out hypothesis (Narloch, Pascual, and Drucker 2012; Kits, Adamowicz, and Boxall 2014; Midler et al. 2015; Chervier, Le Velly, and Ezzine-de-Blas 2019), others studies do not (Travers et al. 2011; Salk, Lopez, and

<sup>☆</sup> This article is part of a special issue entitled: 'Incentives and motivations' published in Ecosystem Services.

\* Corresponding author.

E-mail addresses: [tobias.vorlauffer@zalf.de](mailto:tobias.vorlauffer@zalf.de) (T. Vorlauffer), [steimanis@staff.uni-marburg.de](mailto:steimanis@staff.uni-marburg.de) (I. Steimanis), [plassenberg@gws-os.com](mailto:plassenberg@gws-os.com) (J. Plassenberg).

Wong 2017; Kaczan, Swallow, and Adamowicz 2019; Handberg and Angelsen 2019; Maca-Millán, Arias-Arévalo, and Restrepo-Plaza 2021; Vorlauffer et al. 2023). More recently, researchers have started to discuss to what extent PES can crowd in non-monetary motivations, leading to stronger conservation outcomes than suggested by the pure economic incentives (both when PES are in place and if terminated) (Andersson et al. 2018; Moros, Vélez, and Corbera 2019; Kerr et al. 2019; Grillos et al. 2019; Lliso et al. 2021; Moros et al. 2023).

This article presents the first meta-analysis of crowding effects of PES. We focus on behavioral crowding because most studies in this area employ lab-in-the-field experiments, a well-established method in experimental economics for examining how individuals respond to incentives in controlled settings. While these experiments enable consistent observation of behavioral responses, they often draw conclusions about underlying motivational shifts but without making the necessary assumptions transparent. In contrast, the few studies that directly investigate motivational crowding in real-world PES contexts rely on quasi-experimental designs (e.g. Agrawal, Chhatre, and Gerber 2015; Chervier, Le Velly, and Ezzine-de-Blas 2019; Blanco et al. 2023), or randomized control trials (Vorlauffer et al. 2023; Grillos et al. 2019). However, these studies lack comparability due to widely differing measures of motivational crowding, varying data collection timelines since PES programs ended. Accordingly, this meta-analysis limits its scope to incentivized lab-in-the-field experiments involving real-world resource users, such as farmers or forest-dependent individuals. A detailed theoretical distinction between behavioral and motivational crowding, along with our approach for identifying behavioral crowding effects in these experiments, is provided in Section 2.

Decisions in these experiments are framed in the context of specific conservation issues that participants are facing in their everyday lives. Additional monetary incentives, typically framed as external payments or bonuses for conservation, are then introduced in the experiments to mimic PES. As stated above, the results of these experimental studies have been mixed. However, it remains challenging to determine whether these mixed results are driven by sampling variation, the sampled population, or differences in the underlying experimental designs. Based on a systematic literature search, we identified sixteen publications that met our pre-defined inclusion criteria, of which we successfully acquired eleven datasets. Obtaining the original, individual-level datasets allows us to a) calculate effect sizes using a consistent statistical approach across publications and b) conduct sub-group analysis (see Section 2 for details).

Experimental designs of these publications can be classified into two types that relate to two archetypical conservation scenarios. First, experiments such as the public good game and common pool resource experiments, resemble a scenario whereby conservation efforts primarily generate positive on-site environmental externalities that benefit the group of resource users. The group faces a social dilemma, where the group can maximize their joint income if they sustainably manage the resource, but individuals have an incentive to free-ride and maximize their individual earnings. Here, individual decisions can be driven by strategic considerations and by the belief in what other resource users do. We refer to these as *strategic experiments*, to convey that in these environments strategic behavior is possible.

Second, other experiments, in contrast, model a scenario whereby conservation behavior generates only positive environmental externalities that benefit off-site users, such as downstream water provisioning. Conservation behavior, in this case, always implies that land users forgo their own income for the benefit of third parties. Individual conservation behavior cannot be driven by strategic considerations as pay-offs are not interdependent, so we refer to these games as *non-strategic experiments*.

In reality, conservation issues are often a combination of the two archetypes. Forest conservation, for example, provides local on-site benefits such as firewood and non-timber products, and regional or even global off-site benefits such as water regulation and carbon sequestration, respectively. By comparing existing evidence from both

experimental types, we are – to our knowledge – the first to investigate whether behavioral crowding effects are more pronounced in one of the two experimental types and hence conservation scenarios.

The results of our meta-analysis suggest that economic incentives are, as indicated by standard economic theory, an effective instrument to foster conservation behavior, especially in settings with positive on-site environmental externalities that benefit local resource users. At the same time, we find no support that PES leads to behavioral crowding-out once they have been terminated.

## 2. Literature review and hypotheses

### 2.1. Behavioral and Motivational Crowding

Crowding effects have been explored in psychology and economics through distinct theoretical lenses, linking motivations to behaviour (Festé and Garrouste 2015). In psychology, the most prominent theory emphasizes intrinsic motivations, which are associated with the enjoyment and interest derived from engaging in specific task (Ryan and Deci 2000). Self-Determination Theory posits that experiences of autonomy, competence, and social connectedness enhance intrinsic motivation. Conversely, external interventions—such as positive or negative incentives—that undermine these feelings can diminish intrinsic motivation (Ezzine-de-Blas, Corbera, and Lapeyre 2019; Ryan and Deci 2000). In economics, crowding-out encompasses a wider range of mechanisms, including shifts in decision-making frames, long-term changes in values or personal norms, and concerns about self-image (Rode, Gómez-Baggethun, and Krause 2015; Bowles and Polanía-Reyes 2012; Bénabou and Tirole 2003; 2006; Frey and Jegen 2001; Ariely, Bracha, and Meier 2009). Motivational crowding thus encompasses a broader spectrum of behavioural drivers beyond the intrinsic motivations defined by Ryan and Deci (2000). Adopting this broader perspective beyond intrinsic motivations, we refer to these behavioural drivers as *non-monetary motivations*.

When focusing on crowding effects from a conceptual and theoretical perspective, it is essential to differentiate between behavioral and motivational crowding. *Motivational crowding* occurs when economic incentives affect the underlying non-monetary motivations to engage in the targeted behavior. This can, but does not necessarily, lead to *behavioral crowding*, resulting in an observable change in the targeted behavior itself. The connection between motivation and behavior is, of course, affected by various other factors such as market incentives and constraints, so both types of crowding effects are not necessarily identical.

While in place, PES aim to stipulate the targeted behavior relative to a scenario without PES, which is also known as *additionality*. Whether PES generate additional conservation depends on the pure incentive effect induced by the payments, and motivational crowding effects. In addition, prior pre-incentive behavior (which in turn can be thought of as a function of pre-existing economic incentives and non-monetary motivations) can be decisive, in particular, if land users already engage in the maximum possible conservation (i.e. ceiling effects). We will refer in the remaining paper to the overall effect of PES while in place as incentive effect. Measuring behavioral crowding effects when economic incentives are in place is challenging, as they can only be observed if the net effect of the pure incentives and motivational crowding is zero or negative (i.e., the opposite direction targeted by the incentive). In this case, the positive incentive effect cannot offset the negative effect of reduced non-monetary motivations.

Once the PES have been terminated, motivational crowding effects can play a major role in determining the *permanence* of conservation benefits. In the absence of the pure incentive effects induced by the payments, a crowding-out of motivations can lead to conservation behavior that is below pre-PES levels. This would offset the positive conservation impact generated while PES were in place over time and lead to a negative conservation impact in the long term. Vice versa, it is

also possible to think of motivational crowding-in that leads to a permanent increase in the targeted behavior, even though payments stopped. When observing behavioral crowding effects post-PES, one may however not induce that motivational crowding is at play. PES may have also affected other decision parameters, for example relaxed former capital constraints that could allow land-users to acquire and clear more forestland for farming).

PES have been studied both in terms of behavioral and motivational crowding. In the latter case, empirical studies have focused on different types of motivations (e.g. Moros, Vélez, and Corbera 2019; Blanco et al. 2023), values (Grillos et al. 2019) or beliefs (e.g. Chervier, Le Velly, and Ezzine-de-Blas 2019). However, most studies in this field have implemented lab-in-the-field experiments that measure behavioral crowding (one exception is Moros et al. (2019)). These LitF either explicitly or implicitly assume that any behavioral crowding effects can be attributed to a shift in the underlying motivations, as other factors of the decision environment can be held constantly by the experimenter. In the following section, we therefore discuss the approach and assumptions of LitF experiments for identifying motivation crowding effects.

### 2.2. Baseline experiments

As discussed earlier, LitF experiments have primarily relied on two types of baseline experiments. The first type involves *strategic experiments*, such as public goods or common pool resource experiments, which are interactive and feature interdependent payoffs among participants. A common example is the standard linear public goods experiment (see Equation 1), where the player  $j$ 's payoff in round  $t$  is a function of the initial endowment  $e_{jt}$ , her contribution to the public good  $x_{jt}$  and the individual return from the public goods. With  $0 < \alpha < 1$  and  $n^*\alpha > 1$ , participants face an incentive to free ride on the contributions of others, while it is socially optimal for all players to contribute their entire endowment to the public good (Andreoni and Croson 2008).

$$\pi_{jt} = e_{jt} - x_{jt} + \alpha \sum_{i=1}^n x_{it} \tag{1}$$

In this experiment, individuals' contributions create a positive externality for other players, meaning behavior may be partially influenced by pro-social preferences. Furthermore, in multi-round or indefinitely repeated experiments, participants have an incentive to sustain group contributions to the public good at a high level for as long as possible to maximize their own income. As a result, behavior can also be driven by strategic considerations. Experiments have demonstrated that many individuals condition their actions on their beliefs about the expected behavior of others (Fischbacher, Gächter, and Fehr 2001; Fischbacher and Gächter 2010).

Second, *non-strategic experiments*, such as the dictator game, involve payoffs that depend entirely on an individual's own decisions (see Equation 2).

$$\pi_{jt} = e_{jt} - x_{jt} \tag{2}$$

In this setting, individuals are endowed with resources  $e_{jt}$  and decide by themselves how much to share with the recipient  $x_{jt}$ . Consequently, behavior is driven solely by either pure or impure altruistic preferences (Andreoni 1989; Ottoni-Wilhelm, Vesterlund, and Xie 2017).

### 2.3. Identifying crowding effects with experiments

Regardless of the underlying baseline experiment, LitF experiments employ various strategies to estimate the crowding effects of PES. Table 1 summarizes these strategies and their underlying assumptions. Importantly, they allow to identify behavioral crowding effects but do not directly capture motivational crowding.

Early experiments primarily focused on the net effects of incentives while they were still in place, comparing PES to other conservation

**Table 1**  
Overview of different crowding identification strategies.

Stage	Design	Estimate	Assumption
PES	Within Subject	Net- effect (incentive + crowding).	No change in behavior in the absence of PES.
	Between Subject	Evidence for crowding-out if net-effect $\leq 0$ .	Control group valid counterfactual for PES treatment group.
Post-PES	Within Subject	Behavioral crowding effect after PES removal.	No change in behavior in the absence of PES.
	Within-Between Design (Diff-in-Diff)		Parallel trends assumption

policies (e.g. Vollan 2008; Travers et al. 2011; Narloch, Pascual, and Drucker 2012; Midler et al. 2015). As noted earlier, these studies can detect crowding effects only when the net effect of PES is zero or negative. This approach limits the ability to identify small or medium-sized crowding effects, as they may be obscured by the direct incentive effect. Most of these studies introduced PES after several rounds of the baseline experiment and did not include a control group (Vollan 2008; Narloch, Pascual, and Drucker 2012; Midler et al. 2015). By comparing pre-PES behavior to behavior during PES, crowding effects are estimated under the assumption that behavior would have remained unchanged in the absence of PES. In contrast, Handberg and Angelsen (2019) implemented a between-subject design, allowing for a comparison between participants exposed to PES and those not exposed (control group). This approach relies on the assumption that the control group serves as a valid counterfactual for what would have been observed in the treatment group without PES.

More recently, experiments have adopted designs incorporating a post-PES stage, in which incentives are removed after a predefined period (e.g. Salk, Lopez, and Wong 2017; Kaczan, Swallow, and Adamowicz 2019; Andersson et al. 2018). Most of these experiments measure behavioral crowding effects by comparing behavior in the post-PES stage—after the removal of incentives—with behavior in the pre-PES stage, before incentives were introduced (Andersson et al. 2018; Salk, Lopez, and Wong 2017; Lliso et al. 2021; Bernal-Escobar et al. 2021; Bernal-Escobar, Engel, and Midler 2021). As experimenters have control over most decisions parameters, they can assure that the budgetary constraints as well as underlying incentives are identical in the pre- and post-PES period. Unlike experiments without a post-PES stage, this approach allows for the isolation of behavioral crowding effects from the incentive effect.

Similar to experiments with a within-subject design and only a PES stage, these approaches rely on the assumption that there would be no change in behavior in the absence of PES. However, this assumption may be overly restrictive, as behavior is likely influenced by factors such as fatigue, learning, or shifts in beliefs unrelated to PES. For instance, in Public Good Games, cooperation has been shown to decline over time (Neugebauer et al. 2009). Additional factors, such as reference-dependent preferences based on expected earnings and fatigue effects, may also play a role, even in non-strategic experiments. To address this limitation, more recent designs have incorporated a control group that is not exposed to PES. This addition enables researchers to account for what might have occurred in the absence of PES (Hönow 2021; Moros et al. 2023; Maca-Millán, Arias-Arévalo, and Restrepo-Plaza 2021; Andersson et al. 2018). By applying difference-in-differences estimations, these designs relax the identification assumption of no time effects, replacing it with the more flexible assumption of parallel trends—where, in the absence of PES, both groups would follow a similar trajectory. Although pure between-subject designs could also be implemented with a post-PES stage, we did not find any examples of this approach in the literature. Similarly, we did not encounter any publications featuring a within-between design with only a PES stage.

Therefore, we refrain from discussing this option in more detail.

Experiments with a post-PES stage often conclude that shifts in motivations occur, based on observed changes in behavior. Since experimenters can control incentives between the pre- and post-PES periods, it is either implicitly or explicitly assumed that changes in behavior can be attributed to shifts in motivations and/or beliefs. Beliefs are particularly relevant in strategic experiments, where payoffs partially depend on the decisions of other participants. In such cases, PES could influence beliefs about the behavior of others, with these beliefs persisting even after the PES have been removed. Similar peer effects shifting beliefs regarding other participants' behavior could be in principle also relevant in non-strategic designs. In the three non-strategic experiments included in the meta-analysis no feedback was given on peers' decisions between rounds, thus excluding any peer effects. While experiments could control for the effect of altered beliefs (ideally by an incentivized belief elicitation in each experimental round), we are not aware of any experiments that do this. Due to the challenges to differentiate between a shift in beliefs and motivations, we solely focus on behavioral crowding for the remainder of the article. In general, we believe that prior LitF experiments have insufficiently discussed this leap from observed behavior to underlying motivation, a point that we explore further in the discussion.

#### 2.4. External and ecological validity

While LitF experiments allow for the estimation of behavioral crowding effects with relatively strong internal validity, they raise the question of how much can be inferred from their findings about crowding effects in real-world PES programs. We identified three crucial assumptions.

First, for meaningful conclusions about motivational crowding due to PES, behavior in LitF experiments must be at least partially driven by the same non-monetary motivations that are susceptible to crowding effects outside the lab. This aligns with the concept of ecological validity, which concerns whether the experimental context sufficiently resembles the real-world context of interest (Roe and Just 2009). LitF experiments typically aim to achieve this by incorporating framing relevant to nature conservation or, in a few cases, by having a direct real-world environmental consequences (Lliso et al. 2021). However, these experiments are incentivized and involve monetary consequences. While this is also true for real-life contexts where land-use decisions affect income, the key question is whether non-monetary motivations are adequately represented. Baseline behaviour is often cited as evidence, arguing that purely selfish individuals lack an inherent incentive to engage in conservation. However, such behaviour could also result from other motivations such as warm glow or experimenter demand effects that may not be present outside the experiment. In strategic experiments, it could additionally stem from social preferences or a desire to foster cooperation, aiming to maximize long-term earnings.

Second, it is necessary to assume that experimental treatments sufficiently resemble real-life PES programs, so that they produce directionally similar crowding effects. However, several factors call this assumption into question and may limit the ecological validity of experimental findings. Experimental setups are typically short-term, introducing PES over only a few consecutive rounds, raising uncertainty about whether these brief treatments can capture the same dynamics as year-long real-world interventions. Additionally, real-life PES programs are often accompanied by awareness-raising campaigns, which can foster crowding-in effects. Furthermore, in experiments, incentives—whether framed as PES or otherwise—are always administered by the experimenters, potentially blurring the external nature of PES incentives relative to other payoffs gained from land-uses. Experimental PES treatments may also trigger responses unlikely to occur in real-life scenarios. For example, framing PES as a bonus or reward for conservation behaviour might heighten experimenter demand effects, especially in controlled lab environments where participants are aware

that their behaviour is closely observed (Zizzo 2010). These demand effects could artificially inflate crowding-in observed in experiments.

Third, there is ideally no self-selection of study participants based on potential crowding effects. Participants in LitF experiments, like those in other research methods involving primary data collection, self-select into studies. This poses a challenge if self-selection is correlated with susceptibility to crowding effects—for example, if individuals prone to crowding-out choose not to participate. When monetary earnings are emphasized, more extrinsically motivated individuals are likely to be recruited. This may limit the inclusion of individuals with high intrinsic motivation, who may be more susceptible to strong crowding-out effects, potentially leading to an underestimation of these effects. This concern relates to the concept of external validity, which addresses the extent to which experimental results can be generalized to a larger population.

We come back to the assumptions in the discussion section, where we outline potential strategies for future research to address these challenges.

#### 2.5. Hypotheses

Based on the existing literature, we formulate three hypotheses to test. The first hypothesis pertains to the period during which PES are still in place, where LitF experiments measure the combined net effect of the incentive and any crowding effects (see Table 1). In line with standard economic theory and the underlying rationale of PES (Engel 2016), we expect that higher economic pay-offs from conservation due to PES will increase conservation levels. Therefore, we do not anticipate crowding-out effects to be strong enough to result in a net decrease in conservation behavior while PES are in place.

**H1:** PES led to an increase in average conservation behavior during the incentive stage.

Given the mixed evidence of prior research on behavioral and motivational crowding once PES have been terminated, we expect PES to lead on average to crowding effects without specifying a direction of the hypothesis.

**H2:** PES lead to a crowding-in/ –out of conservation behavior once incentives have been removed.

On average, H2 may show no net impact due to two opposing effects at the participant level that cancel each other out in aggregate. In line with self-determination theory (Ryan and Deci 2000), we hypothesize that the direction of crowding depends on baseline conservation behavior prior to the introduction of PES. Specifically, we propose that introducing external monetary incentives for individuals with strong internalized (non-monetary) motivation—proxied by their conservation behavior during the initial round of the experiment without PES—will lead to a crowding-out of conservation behavior. Conversely, for individuals with relatively low levels of internalized motivation, we expect the introduction of PES to lead to a crowding-in of conservation behavior.

**H3:** Individuals with relatively high levels of pre-PES conservation behavior are prone to crowding-out, while the opposite is true for individuals with relatively low levels of pre-PES conservation behavior.

In addition to these hypotheses, we conduct *further exploratory analyses* to answer the extent to which the underlying experimental design moderates the effectiveness of PES and crowding effects. In particular, we compare strategic with non-strategic experiments. Strategic experiments feature two key differences. First, individual earnings depend partially on the decisions taken by others. Second, groups can maximize their earnings if they sustain long-term cooperation, which implies high conservation outcomes. In these types of experiments, many individuals condition their behavior on their beliefs about the behavior of others

(Fischbacher, Gächter, and Fehr 2001). PES may help groups to strengthen cooperation as the relative payoffs associated with conservation behavior are increased, thus minimizing the potential loss if others do not conserve as well. Once PES are terminated, the underlying experimental setup may provide sufficient incentives to maintain a stable cooperative equilibrium. This would lead to a crowding-in. Non-strategic designs – as the name suggests – lack this interactive element and can be considered a purer individual-level measure of non-monetary conservation motivations (such as altruistic preferences). Here, the risk of crowding-out may be higher.

### 3. Material and methods

#### 3.1. Systematic literature search

Relevant literature was identified through a systematic literature search on three databases (EconLit, Scopus, Web of Science) in June 2022 and complemented by literature identified by a recent systematic review on crowding effects (Engel et al. 2021) and the authors' expert knowledge. The systematic search process is illustrated in Fig. 1,

including the reasons for excluding 19 studies during the full-text screening. The search strings, as well as the literature list subject to the full-text screening and more detailed exclusion reasons with their relation to the eligibility criteria, are provided in the Supplementary Information (SI), Section F. The following inclusion criteria were used to screen the search results for eligible publications:

1. Peer-reviewed articles, doctoral dissertations or published working papers without any restrictions regarding the publication year.
2. The study must be based on a framed lab-in-the-field experiment, also referred to as framed field experiment (Harrison and List 2004), that is characterized by: a) non-student subjects, in particular, subjects who regularly engage in conservation-related behavior such as farmers, forest users, fishers; b) framed in the context of nature conservation or the conservation of natural resources to adequately capture non-monetary motivations, and c) incentivized, implying that the decisions taken in the experiment at least partially determined monetary pay-offs.
3. The experimental design captures prosocial behavior, in particular, decisions to either share resources with someone else in the

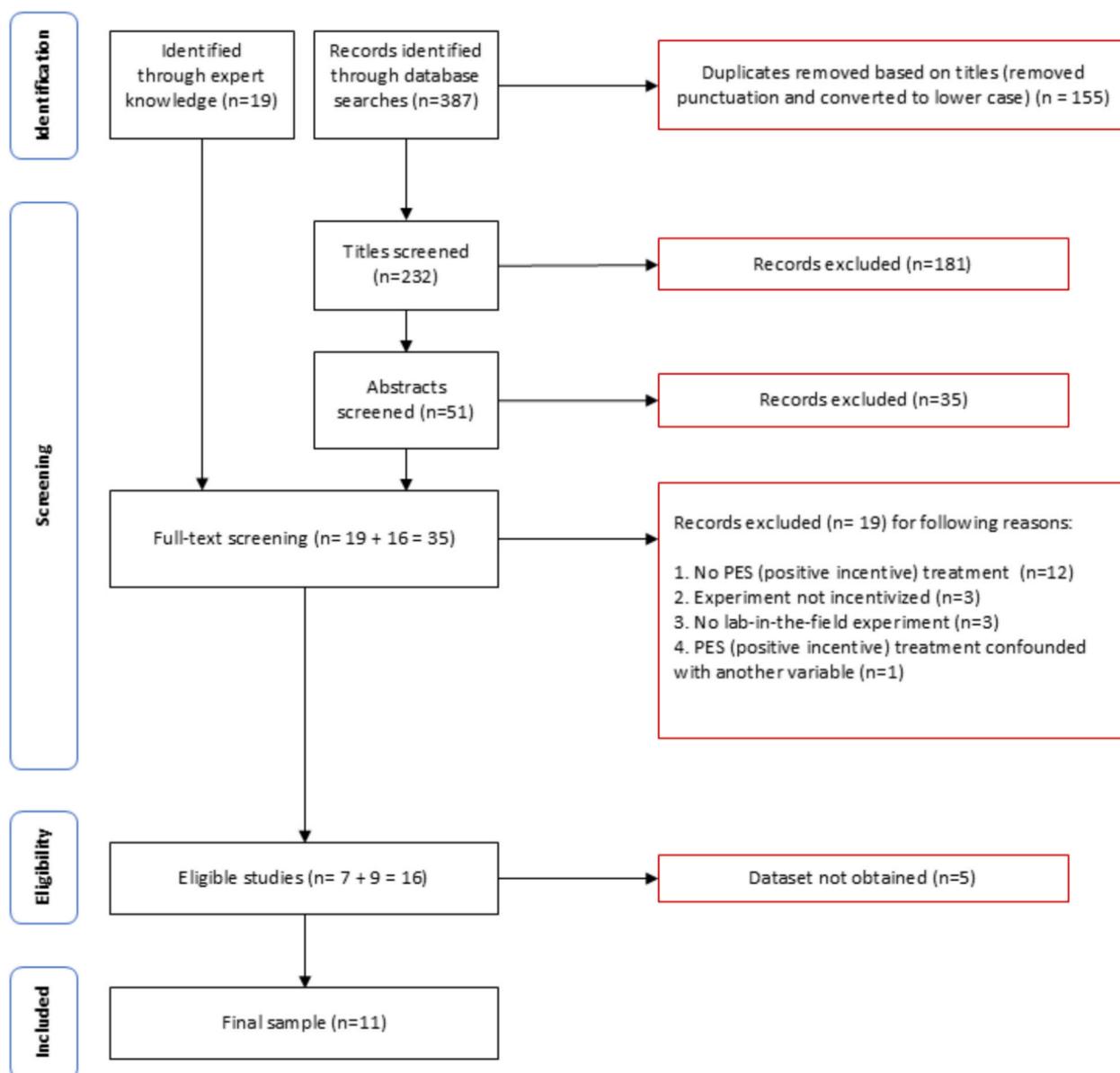


Fig. 1. Prisma flow diagram of the systematic literature search.

experiment or outside of the experiment (a dictator game with people or organizations as recipients), or decisions where groups face a cooperation problem, such as contributing to a public good or joint extraction from a common pool resource. In the latter case, individual payoffs partially depend on the decision taken by other participants.

4. The experimental design includes a counterfactual to which the effects of PES can be estimated. This can be achieved by comparing behavior a) of different participants exposed to PES or not (between-subject design), b) of the same participants over time (within-subject design), or c) a combination of both (between-within design) (see Table 1 for more details). The incentives mimicking PES do not necessarily need to be labelled as PES but could be – for example – framed as additional payments or bonuses conditional on a specific individual or group conservation target.

In total, 16 eligible publications were identified. With five exceptions (Kaczan, Swallow, and Adamowicz 2019; Midler et al. 2015; Dörschner and Musshoff 2015; Narloch, Pascual, and Drucker 2012; Reichhuber, Camacho, and Requate 2009), we received access to the datasets through publicly available repositories or by contacting the authors. Except one study from Germany (Dörschner and Musshoff 2015), all remaining studies were conducted in low income countries. Four of these studies focus solely on the incentive period and do not include a post-PES period. More study details are provided in the SI, Section G, and we discuss the potential bias introduced by this selection in Section 5.

The eleven included publications and their characteristics are summarized in Table 4. More than half of the studies (6) were conducted in Colombia. The most common experimental designs are strategic experiments, i.e. common pool resource games and threshold public good games. Only three out of eleven publications employ non-strategic dictator games. We excluded a few treatments from four publications as they did not include positive conditional conservation incentives that were introduced in the second stage (see Table 4 for details). Out of the 11 included publications, eight include a post-PES period. Only one studies applies a between-subject design (Handberg and Angelsen 2019), whereas four studies use a between-within design (Moros, Vélez, and Corbera 2019; Moros et al. 2023; Hönow 2021; Maca-Millán, Arias-Arévalo, and Restrepo-Plaza 2021). All remaining studies estimate crowding effects within subjects.

### 3.2. Estimating study effect sizes and conducting the meta-analysis

Access to the raw data allows us to a) estimate the crowding effects with the same econometric specifications across publications and b) analyze sub-group effects. For this approach, we first re-estimated individual effect sizes that are later used as foundation for the meta-analysis. Overall, we estimate 42 unique effect sizes referred to as ‘treatment estimates’. We estimate and report individual effect sizes for each treatment. We also differentiate between sub-samples for three publications: Andersson et al. (2018) and Volla (2008) conducted the experiments in five and two countries respectively, and Llisó et al. (2021) conducted their experiments in three different types of Colombian communities. The primary rationale for reporting the individual, treatment-level effect sizes is to illustrate heterogeneity in crowding effects across treatments and sub-samples. Since treatments are independent and were designed with the expectation to result in different degrees of crowding, estimating solely publication-level effect sizes may obfuscate important insights (Borenstein 2009, 217–23). Interested readers are referred to the results in the SI, Section B.6, where we also report the average effects by publication.

Since the experiments in all publications define unique strategy spaces with maximum and minimum conservation levels, it is straightforward to standardize the behavioral outcomes at the subject and round level. The individual decision  $d_{ik}$  of subject  $i$  in round  $k$  in the raw

datasets were standardized as follows:  $cons_{ik} = \frac{d_{ik} - d_{min}}{d_{max} - d_{min}} * 100$  if a larger  $d$  implies more conservation, and  $cons_{ik} = \frac{(d_{max} - d_{ik})}{d_{max} - d_{min}} * 100$  otherwise. The standardized outcome  $cons_{ik}$  ranges between 0 and 100, with 0 indicating no conservation at all and 100 indicating the maximum possible conservation level. Baseline behavior prior to the introduction of PES across all studies indicates that only a fraction of participants does not engage in any conservation behavior (10.5 %) and full conservation behavior (14.8 %), thus potentially allowing for crowding-out among 89.5 % and crowding-in among 85.2 % of participants. Among the strategic experiments, only one publication (Salk, Lopez, and Wong 2017) includes a social optimum (that maximizes the joint group payoff) that does not coincide with the maximum conservation level. In their design, the social optimum is located at 70 % and 40 % of the maximum conservation effort, depending on the treatment. We do not consider the social optimum as the benchmark against one should measure PES performance. PES that achieve conservation above the social optimum, can be still considered environmentally more effective than PES that reach the social optimum.

All treatment estimates are derived from separate mixed-effects Tobit models, which account for the censored nature of conservation decisions  $cons_{ik}$  (censored between 0 and 100) and the non-independence of the data. Since multiple observations are collected per participant across rounds, decisions are correlated within individuals. Additionally, strategic experiments involve group interactions, necessitating the inclusion of nested random effects to account for the non-independence of observations within groups. Specifically, decisions are nested within individuals and groups (Rabe-Hesketh, Skrondal, and Pickles 2005) (by applying the STATA command “metobit conservation PES (or post-PES) || group: || individual, ll(0) ul(100) iterate(500) vce(robust)”). Heteroskedasticity-robust standard errors were employed in all estimation processes.

Depending on the identification strategy and data structure of the original studies (Table 1), we applied different approaches to estimate individual treatment effects. Separate estimations were conducted to measure the incentive effect while PES were in place (H1) and the crowding effect after PES were removed (H2, H3). For within-subject designs, we included data from the pre-PES rounds and the relevant rounds either while PES were in place (H1) or after they were removed (H2, H3). The model specifications included a dummy variable indicating whether the conservation decision  $cons_{ik}$  of individual  $i$  in round  $k$  occurred while PES were in place (H1) or after their removal (H2, H3), thereby capturing the average incentive or crowding effect.

Between-subject estimations were employed for only two publications. Handberg and Angelsen (2019) do not include a pre-PES period, but instead estimate PES effects relative to a control treatment. We therefore combined the three different PES treatments and estimated the overall effect of PES relative to the control treatment. The original paper finds that higher incentives are more successful in increasing conservation. Estimating separate effect sizes for the different PES treatments would have involved repeated use of the control group, resulting in dependent effect sizes that could not be treated as independent (Borenstein 2009, 239–42). A special case is Anderson et al. (2018), who included a pre-PES period and introduced three treatments in a second period—communication, PES, and PES + communication—before reverting to the baseline conditions in a third period. For the PES treatment, we used a within-subject estimation strategy. However, this approach could not be applied to the PES + communication treatment because the effects of communication and PES are perfectly confounded. Consequently, we applied a between-subject estimation, using the communication treatment as the control condition. For each estimation, we included the respective control and treatment groups. The treatment effects were then estimated by incorporating a treatment-level indicator variable into the regression specification.

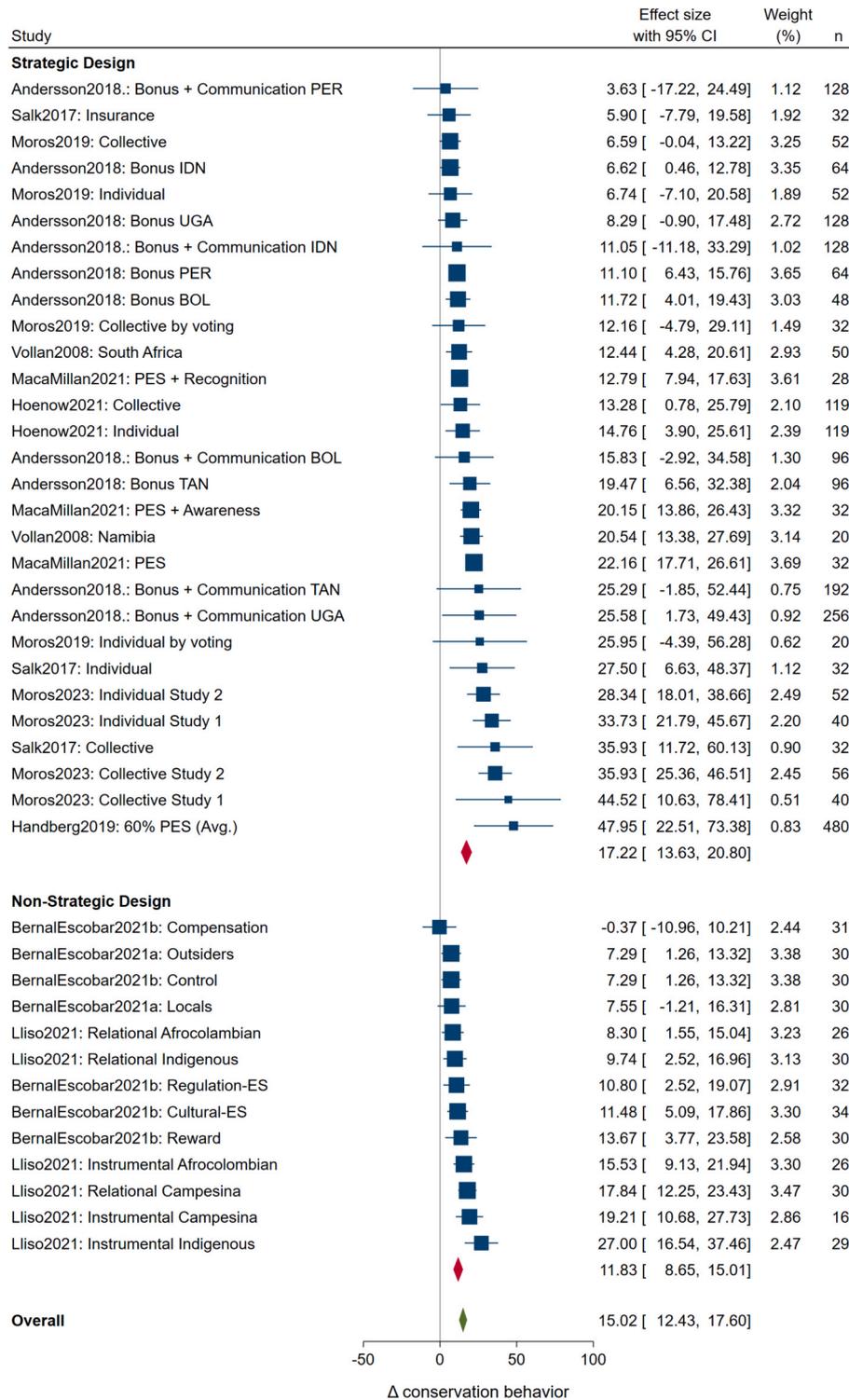
For testing H3, we applied the same approach as for H2 but conducted a sample split and computed separate treatment estimates. The

sample was split by respondents below and above the median of the first round at the publication level.

These individual treatment estimates were then used for the meta-analysis. We report random effects meta-analyses that account for heterogeneity in study populations and experimental designs that likely generate heterogeneous treatment effects (Borenstein 2009, 83–84). Meta-analytic results based on standardized within-group treatment

effects based on pre- and post-PES averages are reported in SI, Section B.3 (Harrer et al. 2021, 68).

Additional robustness checks with leave-one-out meta-analytic estimations, as well as pooled data analysis (where all observations are weighted equally), are provided in SI, Section B.4 and B.2. In order to analyze potential publication biases and correct for them, we implement the funnel-plot based, FAT-PET-PEESE meta-regression procedure



**Fig. 2. Forest plot of incentive effects of PES by experimental design.** The blue squares illustrate the point estimates of the individual effect sizes, with the bars showing the corresponding 95% confidence intervals (CIs). The red diamonds illustrate the average incentive effects by experimental design, while the green diamond illustrates the average incentive effects across all designs. The width of the diamonds illustrates the 95% CIs. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

proposed by Stanley and Doucouliagos (2014). This technique allows to control for a publication bias based on the effect size (larger effects are more likely to get published) and accounts for p-hacking (Irsova et al. 2023).

### 4. Results

#### 4.1. Hypothesis 1

PES work while in place and do not lead to a crowding-out that offsets the incentive effect. On average, PES lead to a  $\theta = 15.02$  percentage point (pp), 95 % CI [12.43, 17.6],  $z = 11.38$ ,  $p < 0.01$ , increase in conservation behavior, thus supporting H1. The individual treatment

estimates are illustrated in Fig. 2 separately for strategic and non-strategic designs. Overall, there is substantial heterogeneity in observed treatment estimates. Only one out of 42 estimates is negative and shows a close to zero effect of PES. The 95 % confidence intervals of eleven additional estimates include the null effect, while the remaining estimates find statistically significant positive incentive effects.

Cochrane’s Q-test rejects the null hypothesis that heterogeneity in estimated effects is occurring only due to sampling variation  $Q(41) = 122.68$ ,  $p < 0.01$ . Most of the variation in the estimated treatment effects is due to between-estimate heterogeneity rather than the sampling error ( $I^2 = 69.45\%$ ). Pooled analyses that account for the non-independence of observations (within participants, experimental groups, and treatments) and uses equal weights are provided in Supplementary

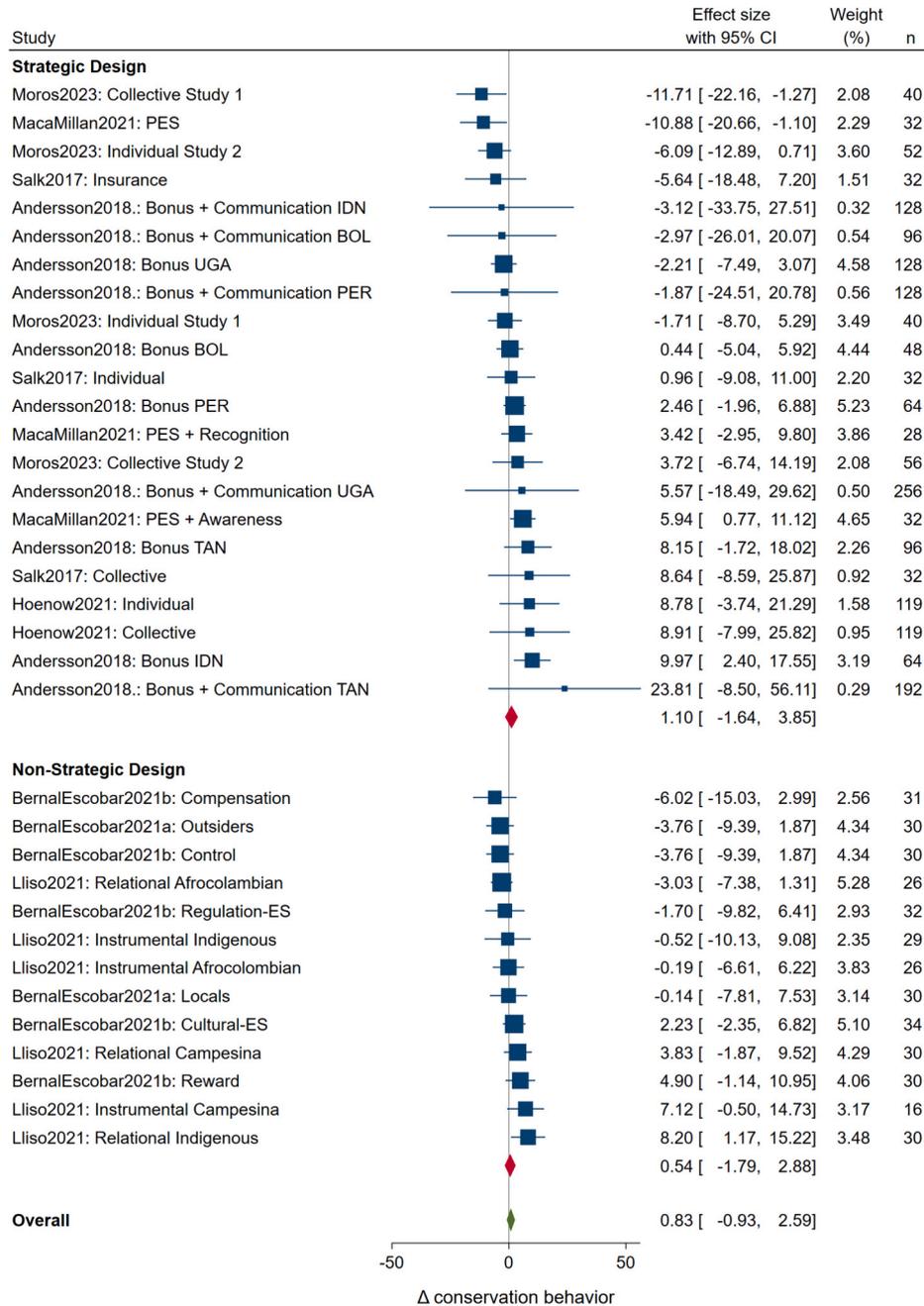


Fig. 3. Forest plot of crowding effects after PES removal by experimental design. The blue squares illustrate the point estimates of the individual effect sizes, with the bars showing the corresponding 95% confidence intervals (CIs). The red diamonds illustrate the average incentive effects by experimental design, while the green diamond illustrates the average incentive effects across all designs. The width of the diamonds illustrates the 95% CIs. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Information (SI), Section B.2, Table S1-S3. Here, PES lead, on average, to an increase in conservation effort of 15.4 – 15.6 pp,  $p < 0.01$ . Overall, PES are successful in increasing conservation levels, implying that the incentive effect must outweigh any crowding-out effect, if present at all. Further analysis finds that incentive effects are significantly different between publications, possibly because of different experimental designs or populations more receptive for PES (see SI, Fig. S19).

Following, Stanley et al. (2022), we also estimated the overall effect size using unrestricted weighted least squares (UWLS, which has been shown to be superior in terms of both bias and type-I error rate compared to conventional random effects meta-analysis, particularly when heterogeneity in effect sizes is correlated with standard errors. Following the recommendation of Irsova et al. (2023), we use wild-bootstrapped standard errors clustered at the publications level to account for the relatively small number of publications (Cameron and Miller 2015). The overall PES effect remains positive, and statistically significant with 14.09 pp., 95 % CI [10.27,17.92],  $p < 0.01$ , without, and 8.13 pp, 95 % CI [2.3,13.95],  $p < 0.05$  with publication level fixed effects respectively (see Table S12).

#### 4.2. Hypothesis 2

The second hypothesis focuses on post-PES crowding effects and we thus limit the sample to studies that include a post-PES period (the number of effect sizes decreases from  $n = 42$  to  $n = 35$ ). The results of the meta-analysis for crowding effects on conservation behavior—once incentives are terminated—are shown in Fig. 3. The overall meta-analytic effect size is  $\theta = 0.83$  pp, 95 % CI [-0.93,2.59],  $z = 0.93$ ,  $p = 0.35$ , implying a tendency that PES lead to a crowding-in of conservation behavior that is, however, not statistically significantly different from zero.

The heterogeneity between treatment estimates is large,  $Q(34) = 57.22$ ,  $p < 0.01$ ,  $I^2 = 42.27$  %. Individual crowding effects range between  $\theta_{i,min} = -11.71$  pp (crowding-out) to  $\theta_{i,max} = 23.81$  pp (crowding-in). Overall, 17 estimates report crowding-out and 18 crowding-in respectively, of which only 2 and 3 are statistically significant at the 5 % level. The observed heterogeneity stems not only from study populations or underlying designs of the baseline experiment, as some publications report opposing crowding effects for different treatment variations (e.g. Salk, Lopez, and Wong 2017; Maca-Millán, Arias-Arévalo, and Restrepo-Plaza 2021).

As a robustness check, we provide results of the pooled data analysis in SI, Section B.2, Table S4-S6. Here, we observe across different econometric specifications a statistically insignificant crowding-in effect between 1.52 and 1.64 pp,  $p > 0.1$ . Overall, our results do not support the crowding-out hypotheses (H2). The UWLS estimation of the treatment effects with clustered SEs at the publication level produces a crowding estimate of 0.79 pp., 95 % CI [-0.87,2.45] without and -0.56 pp, 95 % CI [-11.81,10.69] with publication fixed effects (see Table S13). Further analysis finds that crowding effects are—in contrast to the incentive effects— not significantly different between publications (see SI, Fig. S20).

#### 4.3. Hypothesis 3

To assess whether crowding effects are different for participants with high and low pre-PES levels of conservation behavior, we split the sample based on the median behavior in the first round at the publication level. Fig. 4 illustrates the average crowding effects for these two sub-groups. Overall, we observe a crowding-out of -2.54 pp, 95 % CI [-5.64,0.57] among individuals with above median conservation behavior, while we observe a crowding-in of 4.29 pp, 95 % CI [2.00,6.58] for below median participants. The individual effect sizes are illustrated in SI, Fig. S4. The differences between sub-groups are particularly pronounced for strategic designs and the below median sub-group (possibly because of shifted beliefs based on the feedback on the

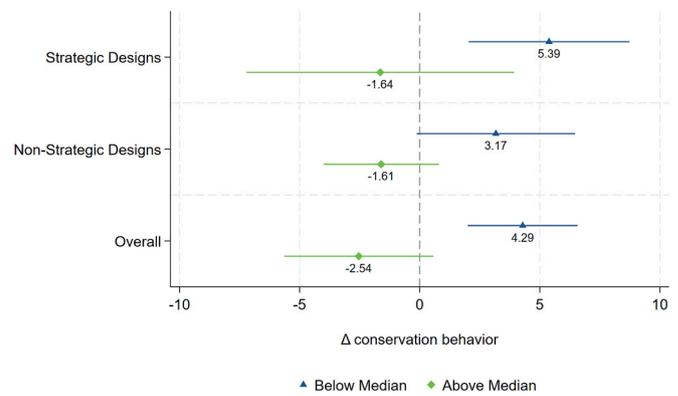


Fig. 4. Average meta-analytic crowding effects by first-round contribution behavior and experimental design.

other participants' behavior), even though the differences between the two experimental designs are not statistically significant ( $Q_b(1) = 0.85$ ,  $p = 0.36$ ).

#### 4.4. Are time trends relevant?

As a robustness check, we restrict our dataset to the participant-level data from four publications that included a control group, which allows for the cleanest measurement of crowding effects as they allow us to account for potential rounds effects. The control groups provide a suitable counterfactual to account for what would have been observed in the absence of PES. Three of these studies included a post-PES stage (Hönöw 2021; Maca-Millán, Arias-Arévalo, and Restrepo-Plaza 2021; Moros et al. 2023), while Moros et al. (2019) only included a PES stage. Table 2, column 1 summarizes the results of a Difference-in-Difference analysis using the pooled respondent-level data from these studies.

We find robust evidence that in the absence of PES conservation behavior already varies over time. In particular, we observe a negative time trend, as conservation behavior decreases in the control groups between the pre-PES and PES stage by 5.36 pp,  $p < 0.05$ . A similar time trend can be observed for the control groups between the pre- and post-PES rounds, even though the effect is not statistically significant. These results suggest that the presented meta-analyses (Fig. 2 and Fig. 3) underestimate the incentive and crowding-in effect, as the analysis focuses on the change over time within individuals and does not account for the overall decline in conservation in the absence of PES. Compared to the meta-analysis based on all publications, we find a relatively stronger increase in conservation due to PES of 25.91 pp,  $p < 0.01$ . We also observe a larger crowding-in effect with a point estimate of 2.99 pp, but with a relatively large CI [-6.96, 12.93]. This robustness check hence corroborates the findings of the main meta-analysis that a crowding-out effect cannot be supported by the available data.

Diverging results between this additional analysis and the main

Table 2  
Incentive and crowding effects estimates for pooled analysis of publications with a control group as difference-in-difference specification. Full model results are reported in Table S7.

	Overall	Below Median	Above Median
Time Trend	-5.36**	-3.14	-10.24***
Intra – Pre	[-10.04,-0.67]	[-7.44,1.17]	[-16.73,-3.76]
Time Trend	-6.65	-2.87	-15.75*
Post – Pre	[-20.10,6.81]	[-15.12,9.38]	[-32.15,0.64]
Incentive effect	25.91***	31.57***	19.27***
	[8.48,43.35]	[13.12,50.03]	[5.96,32.59]
Crowding effect	2.99	4.69	4.29
	[-6.96,12.93]	[-2.60,11.97]	[-8.00,16.59]

95 % confidence intervals in brackets.

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

results may originate from the inclusion of fewer and systematically different studies and not from the different analytical approaches. In SI, Section B.1, we present the main meta-analytic results differentiated by studies that include and do not include a control. Incentives are significantly more effective for the studies with a control, potentially because all four studies have a strategic design where PES are more effective (see Fig. 2). More importantly, we do not find any systematic differences for the crowding effects post-PES.

Controlling for potential time trends through a separate control group is especially relevant for the sub-group analysis that differentiates between high and low conservation levels. For example, in strategic settings, individuals with above median conservation behavior may decrease conservation behavior in the absence of PES after observing lower conservation among fellow participants. The second and third columns in Table 2 provide estimates for the time trends, incentive, and crowding effects for the two sub-groups. Only in the sub-group with above median conservation levels, we observe decreasing conservation behavior in the PES (−10.24 pp,  $p < 0.01$ ) and post-PES stage (−15.75 pp,  $p < 0.1$ ) relative to the pre-PES stage. Overall, PES are more effective in raising conservation behavior in the group with below median conservation levels (31.57 vs 19.27 pp). Surprisingly, we find in both sub-groups indications for similar crowding-in effects of 4.69 pp and 4.29 pp, respectively. But both effects are not statistically significant with wide CIs.

While most studies in our sample do not include a control group, we provide additional robustness checks for those employing a between-subject design. The main analysis presented in Fig. 3 assumes that, in the absence of PES, average conservation behavior would have remained stable relative to the pre-PES period. However, a decline in cooperation is frequently observed in public goods experiments—a pattern that is also reflected in the pre-PES periods of our dataset (see Fig. S21). As a robustness check, we therefore restrict the pre-PES period to only the final two rounds, excluding earlier rounds that could otherwise inflate counterfactual conservation levels. An alternative approach would be to extrapolate the pre-PES time trend as a counterfactual for the PES and post-PES periods. However, the two studies with dedicated control groups (Moros et al. 2023; Maca Millan et al. 2021) indicate that pre-PES trends do not simply continue into subsequent periods (see Fig. S21). For this reason, our robustness check uses only the final two rounds of the pre-PES period to construct the counterfactual.

Under this alternative specification, we find a statistically significant crowding-in effect of 2.66 pp, CI [0.48, 4.84],  $p = 0.02$  (see Figs. S22 and S23, and SI-B.7 for further details). Nonetheless, we maintain a preference for our primary analytical approach, as it is more conservative in detecting crowding-out effects—the dominant concern in the context of PES—by potentially inflating the assumed counterfactual conservation levels.

#### 4.5. Do experimental design features moderate incentive and crowding effects?

We find evidence for systematic differences in incentive effects of PES between strategic and non-strategic designs. The incentive effect of PES on conservation behavior are stronger for strategic ( $\theta = 17.22$  pp, 95 % CI [13.63, 20.80],  $z = 9.41$ ,  $p < 0.01$ ) than non-strategic designs ( $\theta = 11.83$ , 95 % CI [8.65, 15.01],  $z = 7.30$ ,  $p < 0.01$ ), difference  $Q_b(1) = 4.85$ ,  $p = 0.03$ , see also Fig. 2.

Crowding effects are similar for strategic ( $\theta = 1.1$ , 95 % CI [-1.64, 3.85],  $z = 0.79$ ,  $p = 0.43$ ) and non-strategic designs ( $\theta = 0.54$ , 95 % CI [-1.79, 2.88],  $z = 0.46$ ,  $p = 0.65$ ). These estimates are not significantly different from each other ( $Q_b(1) = 0.09$ ,  $p = 0.76$ ). However, the bandwidth of observed crowding effects is larger for strategic than non-strategic designs. Possibly because group interaction leads to more severe outcomes through feedback mechanisms and path dependencies in groups. These results suggest that the underlying game structures do not

explain differences in observed crowding effects between estimates.

The supplementary information provides additional meta-regression results to examine whether other experimental design characteristics are correlated with observed incentive and crowding effects. While the experimental approach is well-suited to study how different PES design variations may moderate crowding effects, the estimated results of the meta-regression should be interpreted with caution, as they do not represent causal effects or robust empirical evidence. Design variations were not randomly assigned across studies and are therefore correlated with each other. Moreover, the variation in certain design features is relatively limited, and the list of potential omitted variables is extensive. Again, the incentive effect is more pronounced for strategic than non-strategic designs when controlling for other design dimensions like priming prior to the introduction of PES, collective instead of individual conditionality for PES, and the existence of a threshold to receive payments, as well as incentive levels (see Table S8). We do not find that any design feature consistently correlates crowding effects across different model specifications (see Table S9).

Again, the incentive effect is more pronounced for strategic than non-strategic designs when controlling for other design dimensions like priming prior to the introduction of PES, collective instead of individual conditionality for PES, and the existence of a threshold to receive payments, as well as incentive levels (see Table S8). We do not find that any design feature consistently moderates crowding effects across different model specifications (see Table S9).

#### 4.6. Publication Bias and Corrections

Fig. 5 illustrates the contour-enhanced funnel plots for the incentive and crowding effects. Asymmetries in funnel plots can be indicative of a publication bias, in particular small-study effects. In the lower left side of Panel A, one can observe slight asymmetries. For a formal test and correction, we follow the FAT-PET-PEESE procedure suggested by Stanley and Doucouliagos (2014). The results are summarized in Table 3. Overall, we find indications for a publication bias with regard to the incentive effect (see Table S10), mirroring the visual inspection of the funnel plot. The bias corrected average incentive effect of PES is smaller with 8.65 pp, 95 % CI [1.51, 15.80],  $p < 0.05$ , than the uncorrected effect of 15.02 pp reported above. The testing procedure does not find indications for a publication bias with regard to the reported crowding effects, which is also reflected by the mixed findings of crowding-in and -out in published research (see also Table S11).

### 5. Discussion

This present meta-analysis yields three main findings: First, economic incentives are, on average, effective in promoting conservation behavior. There is no evidence that motivational crowding can undermine conservation behavior to such an extent that the net-effect of introducing PES is negative, which would render PES non-additional. Overall, the incentives are more effective among individuals with relatively low pre-PES conservation behavior and in strategic settings where conservation generates positive on-site externalities for other resource users. These results suggest that PES can deliver on their conservation promise while they are in place, as corroborated as well by impact assessments of actual PES programs (Snijlsveit et al. 2019; Wunder et al. 2020). While we find indications of a publication bias, favoring publications that report positive incentive effect, the incentive effect is overall robust to correction methods.

However, these findings on the overall effectiveness of incentive-based instruments should caution researchers and practitioners against oversimplifying complex realities or viewing policy instruments like PES as a panacea for achieving more sustainable land-use practices. In recent years, there has been growing recognition that communities worldwide assign diverse values to nature—values that cannot and should not always be reduced to monetary terms, which are often considered a

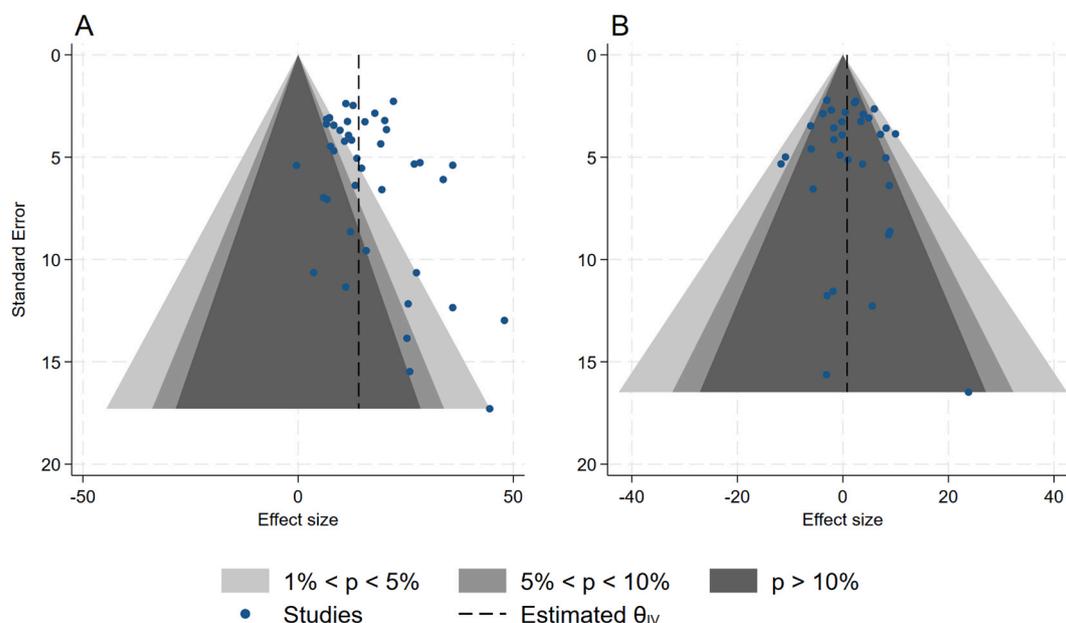


Fig. 5. Contour-enhanced funnel plot for incentive (A) and crowding effects (B).

Table 3

Summary of estimated effect sizes with 95% Confidence Intervals in brackets based on different estimation procedures, and publication bias (PB) diagnosis and correction.

Method	Incentive Effect	Evidence for PB	Crowding Effect	Evidence for PB
Random Effects	15.02 [12.43, 17.60]	--	0.83 [-0.93, 2.59]	--
UWLS with publication fixed effects	8.13 [2.30, 13.95]	--	-0.56 [-11.81, 10.69]	--
FAT-PET-PEESE	8.65 [1.51, 15.80]	Yes	-0.98 [-12.68, 14.64]	No

prerequisite for incentive-based approaches (Pascual et al. 2023). Ideally, PES implementers should remain sensitive to local conditions by acknowledging the diversity of values communities place on nature and actively involving them in the design and decision-making processes during implementation (Bremer et al. 2023). Otherwise, PES interventions risk undermining both positive social and environmental outcomes (IPBES 2022, 289–92).

Second, our results highlight that —on average— PES do not lead to behavioral crowding-out of conservation in LitF experiments once the incentives have been removed. Therefore, experimental evidence cannot substantiate concerns about the unintended impact of PES on conservation in the long run. This finding is also supported when focusing only on a subset of studies with the most rigorous experimental designs that allow to control for time trends.

Notwithstanding the potential limitations of using lab-in-the-field experiments for measuring crowding effects that we touched upon in Section 2 and discuss in more detail below, similar findings are emerging from the literature that evaluates the long-term impact of terminated or suspended real-world PES by focusing on specific measures for motivational crowding (Vorlauffer et al. 2023; Blanco et al. 2023; Grillos et al. 2019). In light of this evidence, it is relevant to revisit the quantitative, observational studies that found evidence for motivational crowding-out. Chervier et al. (2019) elicited perceived benefits of forest conservation among communities that received PES in Cambodia. They match 11 PES villages with 5 control villages based on administrative data

(forest cover, population density and distance to roads), and conclude that PES recipients report money-related benefits of conservation more often and subsistence-related benefits less often than respondents from control communities. While the paper raises some methodological concerns (small sample, matching on a few variables), the paper elicits perceived benefits of forest conservation in general and not motivations for conservation while PES was in place or after being terminated.

Rico García-Amado et al. (2013) conducted surveys in eleven communities within a protected area in Mexico, in which some people received government-funded PES. They find that PES participation is positively correlated with the perception that monetary benefits accrue from the protected area and negatively correlated with the perception that the protected area generates intrinsic benefits for nature. Again, here the focus is not on motivations for conservation behavior but rather perceived benefits (specifically of protected area). More importantly, the paper cannot address the self-selection of communities and individuals in PES and thus cannot rule out systematic differences between PES participants and non-participants before PES were introduced. Another paper by Agrawal et al. (2015) is widely cited as further support for the crowding-out hypotheses of PES. The actual intervention that the study evaluates can be best described as an integrated development and conservation project, whereby communities and households received in-kind support in exchange for mandatory participation in awareness events and trainings. Based on the information provided in the article, we do not consider the program a PES. Taken together, we believe that the existing, observational quantitative evidence in favor of crowding-out due to PES is rather weak. It remains open whether their findings are genuine crowding-out effects (due to conditions specific to the PES under evaluation) or artefacts due to their research designs. But even if PES crowded-out motivations in these two case studies, crowding-out seems more like an exemption than the norm. From our perspective, the research field should exert more caution in concluding that PES regularly crowd-out recipient’s non-monetary motivations.

Generally, we see value in combining insights from randomized controlled trials and quasi-experiments that aim to estimate the crowding effects of actual PES schemes with more abstract experimental studies. While the former likely have greater ecological (external) validity, experimental approaches offer relatively high internal validity for effect estimates. Moreover, LitF experiments allow for testing the impact of multiple PES design choices with relatively low resource requirements, which is often not feasible with actual PES programs.

**Table 4**  
Characteristics of included publications.

Publication	Country	N <sup>a</sup>	Included Treatments	Treatment Type	PES conditionality level	Threshold for PES	Communication	Social Optima w/o PES	Incentive <sup>b</sup>	Excluded Treatments	Experiment Type	Rounds per Stage	Stages	Control Group	
Andersson et al. (2018)	Bolivia, Indonesia, Peru, Tanzania, Uganda	144, 192, 192, 288, 384	Bonus	PES	Collective	Yes (50 % with increasing probability the more forest remained)	No	Corner	25 %		PGG framed as forest resource appropriation	8	3	No	
			Bonus + Communication				Yes							Yes	Yes
			(Communication <sup>h</sup> )	--	--	--	Yes		--					--	--
Bernal-Escobar et al. (2021b)	Colombia	60	Outsiders Locals	PES	Individual	No	--	Corner	100 %	--	DG framed as a positive forest externality for downstream water user	1	3	No	
Bernal-Escobar et al. (2021a)	Colombia	157	Control Cultural-ES Regulation-ES Compensation Reward	PES	Individual	No	--	Corner	100 %	--	DG framed as a positive forest externality for downstream water user	1	3	No	
Handberg and Angelsen (2019)	Tanzania	480	Control PES <sup>d</sup>	PES	Individual	No	Yes	Interior	NA	--	CPR game framed as forest harvesting decision	9	1	Yes	
Hönow (2021)	Namibia	238	Individual Reward Collective Reward	PES	Individual Collective	No Yes (50 %)	No	Corner	12.5 % 6.25 %	Control <sup>c</sup> , Individual Fee	CPR game framed as forest-farmland decision	1 <sup>e</sup>	3	Yes	
Lliso et al. (2021)	Colombia (Samples: Afro-Colombians, Campesinos, Indigenous)	157	Instrumental Relational	PES + priming	Individual	No	--	NA <sup>e</sup>	NA <sup>f</sup>	--	DG framed as forest-farmland decision/ real donation to a local forest conservation program	5	3	No	
Maca-Millán, Arias-Arévalo, and Restrepo-Plaza (2021)	Colombia	92	PES PES + EA PES + R	PES PES + priming PES + priming	Individual	No	No	Corner	22.22 %	Control <sup>c</sup>	Threshold PGG framed as forest-farmland decision	5	3	Yes	

(continued on next page)

Table 4 (continued)

Publication	Country	N <sup>a</sup>	Included Treatments	Treatment Type	PES conditionality level	Threshold for PES	Communication	Social Optima w/o PES	Incentive <sup>b</sup>	Excluded Treatments	Experiment Type	Rounds per Stage	Stages	Control Group
Moros, Vélez, and Corbera (2019)	Colombia	156	Individual payment	PES	Individual	Yes (44 %)	No	Corner	22.22 %	Control <sup>c</sup> , Crop-price premium payment <sup>g</sup>	PGG framed as forest-farmland decision	5	2	Yes
			Collective payment		Collective									
			Individual by voting		Individual									
			Collective by voting		Collective									
Moros et al. (2023)	Colombia	188	IND-TR Study 1	PES	Individual	No	No	Corner	56.25 %, 42.19 %	NO PES <sup>c</sup> , IND-PRR, IND-PRW, COL-PRW, COL-PRR	PGG framed as forest-farmland decision	4	3	Yes
			COL-TR Study 2		Collective									
			IND-TR Study 2		Individual									
			COL-TR Study 2		Collective									
Salk (2017)	Lao	96	Individual Payment	PES	Individual	Yes (30 %)	Yes	Interior	5.13 %	--	PGG framed as forest-farmland decision	8	3	No
			Insurance		Individual									
			Group Payment		Collective									
Vollan (2008)	Namibia, South Africa	20, 50	Reward (by voting)	PES	Individual	Yes (75 %) <sup>h</sup>	No	Interior	5.46 % <sup>i</sup>	Penalty Communication	CPR game framed as grazing on common lands	10	2	No

Notes: <sup>a</sup> Total individual participants of included treatments; <sup>b</sup> For strategic designs we calculated the increase in individual earnings due to PES, if groups reach the social optimum. For non-strategic designs, we calculated the relative increase in earnings for conservation actions due to PES; <sup>c</sup> Control Treatments do not introduce positive incentives in stage 2; <sup>d</sup> We merge all three PES treatments to avoid using the same control group multiple times for a comparison; <sup>e</sup> Each stage consists of 4 rounds where subjects took a binary decision between clearing a new forest parcel or not. This is the only publication with such a dichotomous outcome. To apply the same analysis, we sum up the four rounds to one to generate a quasi-continuous outcome between 0 and 4; <sup>f</sup> Without PES conservation behavior is not generating any pay-off. Pay-offs from PES are slightly below average payments for non-conservation behavior (96 %); <sup>g</sup> This treatment models economic incentives through eco-labelling and not through PES as payments are based on the agricultural area; <sup>h</sup> The threshold is defined as 25 % of the maximum number of sheep; <sup>i</sup> Calculated with the expected value of the reward based on a 20 % probability; <sup>j</sup> The communication treatment serves as control for the bonus + communication treatment.

A growing body of studies have also investigated the permanence of PES, which is the long-term provision of environmental services even in the absence of PES (Engel, Pagiola, and Wunder 2008). The permanence is in turn influenced by the impacts while PES are in place (additionality) and the impacts after PES termination. Crowding-out effects can be pivotal when PES stop, as the targeted behavior can fall below pre-PES levels due to the lack of non-monetary conservation motivations and thus offset the conservation gains achieved during the PES-period. Eventually, this may render PES detrimental in the long-term, even though conservation benefits are created during the PES-period. Studies that focused on the permanence have mostly found positive long-lasting impacts beyond the PES-period (Pagiola, Honey-Rosés, and Freire-González 2016; 2020; Rasch et al. 2021; Calle 2020; Hayes et al. 2022; Etchart et al. 2020). Two studies focusing on PES programs that aimed for avoided deforestation report that PES recipients resumed deforestation activities, similar to levels of land users who never received PES. But gains achieved during the PES-period could still be detected in the long-term (World Bank 2018; Kemigisha et al. 2023). While these studies do not directly measure crowding effects, their results in fact indicate that crowding-out effects are not sufficiently strong to compromise the permanence of PES.

Third and finally, we analyzed the extent to which different behavioral crowding effects occur among individuals with relatively high and low pre-PES conservation behavior. While evidence across all included studies suggests a crowding-in effect among individuals with lower initial conservation behavior, this result is not robust when limiting the analysis to studies that control for time trends in conservation behavior in the absence of PES. Moreover, the analyses do not find evidence of crowding-out effects among individuals with high pre-PES conservation behavior.

### 5.1. Limitations and future research avenues

Our meta-analysis re-estimated treatment effects with the original datasets of the included studies, which allows for a standardized analytical strategy and for estimating sub-group treatment effects for participants with initially high and low levels of conservation behaviour. The downside of this approach is a potential selection bias as we could not access datasets of five eligible publications. The findings of these studies are, however, largely in line with our meta-analysis, suggesting that the inclusion of these studies would have unlikely shifted the overall results. All five studies find that PES are effective in raising the stipulated behavior while in place, except specific PES designs such as collective (Kaczan, Swallow, and Adamowicz 2019; Narloch, Pascual, and Drucker 2012) and action-based payments (Dörschner and Musshoff 2015). Moreover, the only studies that includes a post-PES period to measure crowding effects after PES termination, yields qualitatively similar results as our meta-analysis (Kaczan, Swallow, and Adamowicz 2019). Using reported effect sizes instead of relying on the primary datasets is commonly done with meta-analysis but requires that publications reported either standardized effects sizes or sufficient information to calculate them. In the SI, Section G, we provide details on the information that was provided in the five omitted studies. In SI B.3 we report meta-analytic results based on a standardized within-subject effect size measure. To calculate this measure, one requires the pre- and post-incentive means, standard deviations, the correlation between pre- and post-incentive outcomes, and the sample size. None of the studies report standardized effect sizes, nor sufficient information to calculate them.

Lastly, future research should aim to address the limitations of the existing literature and, by extension, this study. Section 2 provided an initial overview of the key assumptions required to draw meaningful conclusions from LitF experiments about the real-world implications of PES and their influence on motivations. Building on this foundation, we extend the discussion to specifically examine the sampled publications, highlighting their limitations and suggesting potential avenues for

future research.

While the experimental method provides a robust framework for estimating well-identified causal effects (Roe and Just 2009), we identify two key concerns regarding the **internal validity** of current experimental studies. **First**, most included publications—except for four—fail to adequately account for the declining trend in conservation behavior over time within experimental settings. This limitation is particularly significant for experimental designs involving strategic interactions, which constitute the majority of studies. In such designs, material gains are interdependent, and groups face social dilemmas. A well-established finding in experimental economics is that cooperation tends to decrease over time in these scenarios (Neugebauer et al. 2009). One mechanism underlying this decline may be peer effects, as subjects typically receive feedback on the behavior of others. A substantial share of individuals are conditional cooperators who adjust their behavior based on the expected actions of others (Fischbacher, Gächter, and Fehr 2001). A shift in beliefs caused by information about peer behavior can therefore lead to behavioral convergence and a decline in cooperation. These dynamics are particularly relevant for subgroup analyses based on pre-PES conservation behavior. Indeed, our analysis finds that individuals with above-median conservation behavior exhibit a significant downward trend in conservation without PES, potentially due to this mechanism, whereas no significant time trend is observed among below-median individuals. To address this issue, experimental designs should prioritize between-subjects designs, which allow researchers to control for time trends by comparing outcomes across treatments rather than within individuals over time. Without accounting for these dynamics, most of the existing experimental literature is likely to systematically overestimate crowding-out effects.

**Second**, many of the current studies are likely underpowered, with sample sizes typically ranging between 30 and 60 observations per treatment. This small sample size, coupled with publication bias, increases the risk of producing inflated treatment effect estimates (Ioannidis, Stanley, and Doucouliagos 2017). Underpowered studies not only threaten the reliability of individual results but also make it difficult to identify meaningful differences across experimental design features and study sites. The diverging results reported in different publications may, therefore, be largely attributable to sampling error, casting doubt on the robustness of conclusions drawn from individual studies. This limitation also extends to systematic reviews, which often place undue emphasis on findings from individual studies. In our view, reviews frequently attempt to explain divergent study results (Rode, Gómez-Baggethun, and Krause 2015; Akers and Yasué 2019; Barton et al. 2022) without adequately accounting for the constraints imposed by small sample sizes and heterogenous experimental designs. Past conceptual papers have outlined compelling arguments what type of PES characteristics may induce specific crowding effects (Ezzine-de-Blas, Corbera, and Lapeyre 2019). But the fragmented nature of implemented treatments and small sample sizes severely limit the conclusion that can be drawn from both individual studies and this meta-analysis. We believe that more sufficiently powered studies with comparable study designs are needed to draw conclusions under what conditions crowding-out and crowding-in effects are to be expected.

In addition to these two concerns, we shift the discussion to the **external and ecological validity** of LitF experiments. While external validity raises the question of whether a specific study's results apply to different people, situations, or periods (Camerer 2015; Al-Ubaydli and List 2015), ecological validity more specifically focuses on whether results observed in the lab can be extrapolated to behavior in the "wild" (Roe and Just 2009).

**First**, the sampling of experimental participants is non-random, which poses a problem if self-selection is correlated with potential crowding effects. All reviewed experiments incentivize behavior in the baseline, potentially attracting extrinsically motivated and/or egoistic respondents who prioritize monetary payoffs. From a methodological standpoint, assessing the problem of self-selection is challenging, as

susceptibility to crowding effects can only be approximated through observable data, such as baseline motivations or behavior. However, such data, representative of the entire population, is typically unavailable. We argue that moving away from baseline experiments tied to monetary payoffs—by, for example, incorporating real-life conservation behavior or real-effort tasks (cf [Loft et al. 2020](#)), could reduce the likelihood of recruiting monetarily motivated individuals, thereby introducing less bias to the sample.

**Second**, such alternative experimental tasks could also increase the credibility of the incentive treatment in the experiments by triggering a response more comparable to real-life PES. Participants would be better able to differentiate between the external incentives provided by the experimenter (mimicking PES) and the monetary or non-monetary benefits inherently associated with the baseline behavior targeted by PES.

**Third**, the current body of studies focuses on a few countries (6 out of 11 studies were conducted in Colombia) which restricts the generalizability of the results to other contexts. We think that the focus on Colombia is predominantly driven by a combination of two factors. On the one hand, PES are widely implemented in Latin America in general and also in Colombia. On the other hand, Colombia hosts a large and active community of experimental economists who study environmental issues. We therefore do not believe that authors strategically focused on Colombia due to the expected findings, which would further amplify concerns about the external validity. Additional analysis provided in SI, B.5 does not find significantly different effects between Colombian and Non-Colombian studies, both in terms of incentive and crowding effects. Nevertheless, the relatively narrow geographical focus of existing studies should be considered a limitation of the current body of evidence.

**Fourth**, and most importantly, most of the current research on crowding effects primarily relies on lab-in-the-field experiments that observe behavioral crowding. By controlling experimental parameters, it is implicitly or explicitly assumed that any observed change in behavior can be attributed to a relevant change in motivations. However, we have concerns about why this may not always be the case, or why shifts in motivations may have limited ecological validity when it comes to real-world PES. Most of the (included) experiments apply strategic designs to measure crowding effects. However, such designs cannot rule out that beliefs regarding the behavior of other subjects are also affected by (temporary) PES. For example, a positive effect of PES on beliefs and a negative effect on motivations (crowding-out) could cancel each other out, resulting in no observable behavioral differences. Such alternative mechanisms may be particularly relevant when differentiating crowding effects based on baseline conservation levels. Individuals with behavior at the extremes (i.e., very high or very low conservation levels) will, in most cases, receive feedback indicating that average peer behavior is less extreme. If these individuals are responsive to such shifts in beliefs —i.e., if they are conditional cooperators— their behavior gravitates toward the average. Such a mechanism can potentially dominate behavioral crowding effects and caution is therefore warranted when interpreting such analyses as evidence for or against motivational crowding. Future experiments should ideally implement non-strategic designs to measure motivational crowding or use strategic designs that allow to control for changes in beliefs. This would help disentangle effects attributable to shifts in beliefs and motivations, and derive more robust findings on motivational crowding.

Moreover, behavioral crowding in LitF may also be an artefact driven by experimenter demand effects, which could be amplified by the PES treatment framing (i.e., highlighting the societal importance of the incentivized behavior) ([Zizzo 2010](#)). Ideally, future experiments should implement designs that can bound potential demand effects ([de Quidt, Haushofer, and Roth 2018](#)) or minimize such effects, for example, by following a strict double-blind protocol ([Vorlauffer 2019](#)).

On a more general level, future research ideally strengthens the empirical link between motivations and behavior both inside and

outside the lab. So far, only a small number of studies have attempted to establish a connection between pro-environmental behavior and the motivations that drive it ([Moros, Vélez, and Corbera 2019](#); [Blanco et al. 2023](#); [Vorlauffer et al. 2023](#)). We consider it imperative, as a first step, to gain a better understanding of how motivations, and other underlying drivers of behavior such as values and beliefs, are related to actual conservation behavior (in terms of effect sizes and explanatory power) in the absence of PES. Studying whether the introduction of monetary incentives then impacts any of these identified drivers, would provide more nuanced insights into the specific mechanisms. For example, measuring motivations before and after the experiment can provide valuable insights into the relationship between behavior and motivations, as well as the impact of incentives on self-stated motivations ([Moros, Vélez, and Corbera 2019](#)). PES may, for example, alter the type of non-monetary motivations land users have, without necessarily manifesting itself at the behavioral level. If carefully designed and rigorously implemented, LitF experiments can meaningfully contribute to this future research.

## 6. Conclusion

In the present meta-analysis, we investigate to what extent lab-in-the-field experiments with real-world resource users provide support for the frequently voiced crowding-out hypothesis of PES ([Rode, Gómez-Baggethun, and Krause 2015](#); [Ezzine-de-Blas, Corbera, and Lapeyre 2019](#)). Gaining access to the datasets of 11 publications, we re-estimated the incentive effect while PES are in place as well as the crowding effect once PES have been removed. While these experiments succeed in fostering conservation behavior through providing additional economic incentives), we do not find any support for behavioral crowding-out once payments stop. These results overall suggest that crowding-effects as a result of PES may be less of a concern than often voiced in the debate about incentive-based conservation policies.

However, methodological concerns regarding the internal and external validity of current experiments raise questions about the broader applicability of these findings. Further studies are needed to build a more robust evidence base, in particular, regarding the question of how non-monetary motivations are related to conservation behavior in labs-in-the-field and outside, whether these results can be replicated in more diverse contexts, and with sufficiently powered rigorous experimental designs that can control for time trends.

## CRedit authorship contribution statement

**Tobias Vorlauffer:** Writing – review & editing, Writing – original draft, Formal analysis, Data curation, Conceptualization. **Ivo Steimanis:** Writing – review & editing, Validation, Conceptualization. **Jan Plasenberg:** Writing – review & editing, Conceptualization.

## Funding

Funding was received from the German Research Council (DFG), Grant-No: 411113521.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgement

We are especially grateful for the support of the authors of the included studies for providing the datasets and supporting information. The feedback of Carl Salk and Björn Vollan on earlier versions of the manuscript is highly appreciated. Finally, we thank the two anonymous

reviewers and the guest editor for their insightful comments and suggestions.

## Appendix A. Supplementary Information

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ecoser.2025.101750>.

## Data availability

The replication package including the analysis scripts for STATA, and the computed effect sizes are publicly available on the Open Science Framework repository (DOI: [10.17605/OSF.IO/UTHN2](https://doi.org/10.17605/OSF.IO/UTHN2)). For intellectual property right reasons, we cannot provide the raw datasets of the included publications.

## References

- Agrawal, A., Chhatre, A., Gerber, E.R., 2015. Motivational Crowding in Sustainable Development Interventions. *Am. Polit. Sci. Rev.* 109 (3), 470–487. <https://doi.org/10.1017/S0003055415000209>.
- Akers, J.F., Yasué, M., 2019. Motivational Crowding in Payments for Ecosystem Service Schemes: A Global Systematic Review. *Conserv. Soc.* 17 (4), 377–389.
- Al-Ubaydli, O., List, J.A., 2015. On the Generalizability of Experimental Results in Economics. In: Fréchet, G.R., Schotter, A. (Eds.), *Handbook of Experimental Economic Methodology*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195328325.003.0022>.
- Andersson, K.P., Cook, N.J., Grillos, T., Lopez, M.C., Salk, C.F., Wright, G.D., Mwangi, E., 2018. Experimental Evidence on Payments for Forest Commons Conservation. *Nat. Sustainability* 1 (3), 128–135. <https://doi.org/10.1038/s41893-018-0034-z>.
- Andreoni, J., 1989. Giving with Impure Altruism: Applications to Charity and Ricardian Equivalence. *J. Polit. Econ.* 97 (6), 1447–1458.
- Andreoni, J., Croson, R., 2008. Chapter 82 Partners versus Strangers: Random Rematching in Public Goods Experiments. In: Plott, C.R., Smith, V.L. (Eds.), *Handbook of Experimental Economics Results*, 1. Elsevier. [https://doi.org/10.1016/S1574-0722\(07\)00082-0](https://doi.org/10.1016/S1574-0722(07)00082-0).
- Angelsen, A., Brockhaus, M., Duchelle, A.E., Larson, A., Martius, C., Sunderlin, W.D., Verchot, L., Wong, G., Wunder, S., 2017. Learning from REDD+: A Response to Fletcher et Al. *Conserv. Biol.* 31 (3), 718–720. <https://doi.org/10.1111/cobi.12933>.
- Ariely, D., Bracha, A., Meier, S., 2009. Doing Good or Doing Well? Image Motivation and Monetary Incentives in Behaving Prosocially. *Am. Econ. Rev.* 99 (1), 544–555. <https://doi.org/10.1257/aer.99.1.544>.
- Barton, David N., Rebecca Chaplin-Kramer, Elena Lazos Chavero, Meine Van Noordwijk, Stefanie Engel, Alexander Girvan, Thomas Hahn, et al. 2022. 'Chapter 4. Value Expression in Decision-Making'. Zenodo. doi: 10.5281/zenodo.7154678.
- Bénabou, R., Tirole, J., 2003. Intrinsic and Extrinsic Motivation. *Rev. Econ. Stud.* 70 (3), 489–520. <https://doi.org/10.1111/1467-937X.00253>.
- Bénabou, R., Tirole, J., 2006. Incentives and Prosocial Behavior. *Am. Econ. Rev.* 96 (5), 1652–1678. <https://doi.org/10.1257/aer.96.5.1652>.
- Bernal-Escobar, Adriana, Stefanie Engel, and Estelle Midler. 2021. 'Beyond a Market Discourse: Is Framing a Solution to Avoid Motivational Crowding-Out in Payments for Ecosystem Services?' SSRN Scholarly Paper 3910112. Rochester, NY: Social Science Research Network. <https://papers.ssrn.com/abstract=3910112>.
- Bernal-Escobar, Adriana, Stefanie Engel, Estelle Midler, and Tobias Vorlauffer. 2021. 'Who Is Benefiting Downstream? Experimental Evidence on the Relevance of Upstream-Downstream Geographic Distance for Water Provision'. SSRN Scholarly Paper. Rochester, NY. doi: 10.2139/ssrn.3911047.
- Blanco, Esther, Lina Moros, Alexander Pfaff, Ivo Steimanis, Maria Alejandra Velez, and Björn Vollen. 2023. 'No Crowding out among Those Terminated from an Ongoing PES Program in Colombia'. *Journal of Environmental Economics and Management*, May, 102826. doi: 10.1016/j.jeem.2023.102826.
- Borenstein, M. (Ed.). 2009. *Introduction to Meta-Analysis*. John Wiley & Sons, Chichester, U.K.
- Bowles, S., Polanía-Reyes, S., 2012. Economic Incentives and Social Preferences: Substitutes or Complements? *J. Econ. Lit.* 50 (2), 368–425. <https://doi.org/10.1257/jel.50.2.368>.
- Bremer, L.L., Nelson, S., Jackson, S., Izquierdo-Tort, S., Lansing, D., Shapiro-Garza, E., Echavarría, M., et al., 2023. Embedding Local Values in Payments for Ecosystem Services for Transformative Change. *Curr. Opin. Environ. Sustain.* 64 (October), 101354. <https://doi.org/10.1016/j.custos.2023.101354>.
- Calle, A., 2020. Can Short-Term Payments for Ecosystem Services Deliver Long-Term Tree Cover Change? *Ecosyst. Serv.* 42 (April), 101084. <https://doi.org/10.1016/j.ecoser.2020.101084>.
- Camerer, C.F., 2015. The Promise and Success of Lab-Field Generalizability in Experimental Economics: A Critical Reply to Levitt and List. In: Fréchet, G.R., Schotter, A. (Eds.), *Handbook of Experimental Economic Methodology*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195328325.003.0016>.
- Cameron, A.C., Miller, D.L., 2015. *A Practitioner's Guide to Cluster-Robust Inference*. *J. Hum. Resour.* 50 (2), 317–372.
- Chan, K.M.A., Anderson, E., Chapman, M., Jespersen, K., Olmsted, P., 2017. Payments for Ecosystem Services: Rife With Problems and Potential—For Transformation Towards Sustainability. *Ecol. Econ.* 140 (October), 110–122. <https://doi.org/10.1016/j.ecolecon.2017.04.029>.
- Chervier, C., Le Velly, G., Ezzine-de-Blas, D., 2019. When the Implementation of Payments for Biodiversity Conservation Leads to Motivation Crowding-out: A Case Study From the Cardamoms Forests, Cambodia. *Ecol. Econ.* 156 (February), 499–510. <https://doi.org/10.1016/j.ecolecon.2017.03.018>.
- Dörschner, T., Musshoff, O., 2015. How Do Incentive-Based Environmental Policies Affect Environment Protection Initiatives of Farmers? An Experimental Economic Analysis Using the Example of Species Richness. *Ecol. Econ.* 114 (June), 90–103. <https://doi.org/10.1016/j.ecolecon.2015.03.013>.
- Engel, S., 2016. The Devil in the Detail: A Practical Guide on Designing Payments for Environmental Services. *International Review of Environmental and Resource Economics* 9 (1–2), 131–177. <https://doi.org/10.1561/101.000000076>.
- Engel, Stefanie, Juan Felipe Ortiz-Riomalo, Julian Rode, and Fabian Thomas. 2021. 'IPBES VA Chapter 4. Systematic Review on Motivational Crowding by Economic Incentives in Conservation Policies', September. doi: 10.5281/zenodo.6502495.
- Engel, S., Pagiola, S., Wunder, S., 2008. Designing Payments for Environmental Services in Theory and Practice: An Overview of the Issues. *Ecological Economics, Payments for Environmental Services in Developing and Developed Countries* 65 (4), 663–674. <https://doi.org/10.1016/j.ecolecon.2008.03.011>.
- Etchart, N., Freire, J.L., Holland, M.B., Jones, K.W., Naughton-Treves, L., 2020. What Happens When the Money Runs out? Forest Outcomes and Equity Concerns Following Ecuador's Suspension of Conservation Payments. *World Dev.* 136 (December), 105124. <https://doi.org/10.1016/j.worlddev.2020.105124>.
- Ezzine-de-Blas, D., Corbera, E., Lapeyre, R., 2019. Payments for Environmental Services and Motivational Crowding: Towards a Conceptual Framework. *Ecol. Econ.* 156 (February), 434–443. <https://doi.org/10.1016/j.ecolecon.2018.07.026>.
- Festré, A., Garrouste, P., 2015. Theory and Evidence in Psychology and Economics About Motivation Crowding Out: A Possible Convergence? *J. Econ. Surv.* 29 (2), 339–356. <https://doi.org/10.1111/joes.12059>.
- Fischbacher, U., Gächter, S., 2010. Social Preferences, Beliefs, and the Dynamics of Free Riding in Public Goods Experiments. *Am. Econ. Rev.* 100 (1), 541–556. <https://doi.org/10.1257/aer.100.1.541>.
- Fischbacher, Urs, Simon Gächter, and Ernst Fehr. 2001. 'Are People Conditionally Cooperative? Evidence from a Public Goods Experiment'. *Economics Letters*, 8.
- Frey, B.S., Jegen, R., 2001. Motivation Crowding Theory. *J. Econ. Surv.* 15 (5), 589–611. <https://doi.org/10.1111/1467-6419.00150>.
- Gneezy, U., Meier, S., Rey-Biel, P., 2011. When and Why Incentives (Don't) Work to Modify Behavior. *J. Econ. Perspect.* 25 (4), 191–210. <https://doi.org/10.1257/jep.25.4.191>.
- Gómez-Baggethun, E., Muradian, R., 2015. In Markets We Trust? Setting the Boundaries of Market-Based Instruments in Ecosystem Services Governance. *Ecol. Econ.* 117 (September), 217–224. <https://doi.org/10.1016/j.ecolecon.2015.03.016>.
- Grillos, T., Bottazzi, P., Crespo, D., Asquith, N., Jones, J.P.G., 2019. In-Kind Conservation Payments Crowd in Environmental Values and Increase Support for Government Intervention: A Randomized Trial in Bolivia. *Ecol. Econ.* 166 (December), 106404. <https://doi.org/10.1016/j.ecolecon.2019.106404>.
- Handberg, Ø.N., Angelsen, A., 2019. Pay Little, Get Little; Pay More, Get a Little More: A Framed Forest Experiment in Tanzania. *Ecol. Econ.* 156 (February), 454–467. <https://doi.org/10.1016/j.ecolecon.2016.09.025>.
- Harrer, M., Cuijpers, P., Furukawa, T., Ebert, D., 2021. *Doing Meta-Analysis with R: A Hands-On Guide*. Routledge, Boca Raton.
- Harrison, G.W., List, J.A., 2004. Field Experiments. *J. Econ. Lit.* 42 (4), 1009–1055. <https://doi.org/10.1257/00220510430004577>.
- Hayes, T., Murtinho, F., Wolff, H., López-Sandoval, M.F., Salazar, J., 2022. Effectiveness of Payment for Ecosystem Services after Loss and Uncertainty of Compensation. *Nat. Sustainability* 5 (1), 81–88. <https://doi.org/10.1038/s41893-021-00804-5>.
- Hönow, Nils Christian. 2021. 'Field Experiments on Cooperation Behavior', June. doi: 10.17192/Z2021.0224.
- Ioannidis, J.P.A., Stanley, T.D., Doucouliagos, H., 2017. The Power of Bias in Economics Research. *Econ. J.* 127 (605), F236–F265. <https://doi.org/10.1111/eoj.12461>.
- Ipbes, 2022. Methodological Assessment of the Diverse Values and Valuation of Nature of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. <https://doi.org/10.5281/ZENODO.6522522>.
- Irsova, Z., Doucouliagos, H., Havranek, T., Stanley, T.D., 2023. Meta-Analysis of Social Science Research: A Practitioner's Guide. *J. Econ. Surv.* <https://doi.org/10.1111/joes.12595>.
- Kaczan, D.J., Swallow, B.M., W. L. (Vic) Adamowicz., 2019. Forest Conservation Policy and Motivational Crowding: Experimental Evidence from Tanzania. *Ecol. Econ.* 156 (February), 444–453. <https://doi.org/10.1016/j.ecolecon.2016.07.002>.
- Kemigisha, E., Babweteera, F., Mugisha, J., Angelsen, A., 2023. Payment for Environmental Services to Reduce Deforestation: Do the Positive Effects Last? *Ecol. Econ.* 209. <https://doi.org/10.1016/j.ecolecon.2023.107840>.
- Kerr, J.M., Bum, T., Lapinski, M.K., Liu, R.W., Zhi, Lu., Zhao, J., 2019. The Effects of Social Norms on Motivation Crowding: Experimental Evidence from the Tibetan Plateau. *Int. J. Commons* 13 (1), 430–454.
- Kits, G.J., Adamowicz, W.L., Boxall, P.C., 2014. Do Conservation Auctions Crowd out Voluntary Environmentally Friendly Activities? *Ecol. Econ.* 105 (September), 118–123. <https://doi.org/10.1016/j.ecolecon.2014.05.014>.
- Kosoy, N., Corbera, E., 2010. Payments for Ecosystem Services as Commodity Fetishism. *Ecological Economics, Special Section - Payments for Environmental Services: Reconciling Theory and Practice* 69 (6), 1228–1236. <https://doi.org/10.1016/j.ecolecon.2009.11.002>.
- Lliso, B., Arias-Arévalo, P., Maca-Millán, S., Engel, S., Pascual, U., 2021. Motivational Crowding Effects in Payments for Ecosystem Services: Exploring the Role of

- Instrumental and Relational Values. *People Nat.* November, pan3.10280. <https://doi.org/10.1002/pan3.10280>.
- Loft, L., Gehrig, S., Salk, C., Rommel, J., 2020. Fair Payments for Effective Environmental Conservation. *Proc. Natl. Acad. Sci.* 117 (25), 14094–14101. <https://doi.org/10.1073/pnas.1919783117>.
- Maca-Millán, S., Arias-Arévalo, P., Restrepo-Plaza, L., 2021. Payment for Ecosystem Services and Motivational Crowding: Experimental Insights Regarding the Integration of Plural Values via Non-Monetary Incentives. *Ecosyst. Serv.* 52 (December), 101375. <https://doi.org/10.1016/j.ecoser.2021.101375>.
- Midler, E., Pascual, U., Drucker, A.G., Narloch, U., Soto, J.L., 2015. Unraveling the Effects of Payments for Ecosystem Services on Motivations for Collective Action. *Ecol. Econ.* 120, 394–405.
- Moros, L., Vélez, M.A., Corbera, E., 2019. Payments for Ecosystem Services and Motivational Crowding in Colombia's Amazon Piedmont. *Ecol. Econ.* 156 (February), 468–488. <https://doi.org/10.1016/j.ecolecon.2017.11.032>.
- Moros, L., Vélez, M.A., Quintero, D., Tobin, D., Pfaff, A., 2023. Temporary PES Do Not Crowd-out and May Crowd-in Lab-in-the-Field Forest Conservation in Colombia. *Ecol. Econ.* 204 (February), 107652. <https://doi.org/10.1016/j.ecolecon.2022.107652>.
- Narloch, U., Pascual, U., Drucker, A.G., 2012. Collective Action Dynamics under External Rewards: Experimental Insights from Andean Farming Communities. *World Dev.* 40 (10), 2096–2107.
- Neugebauer, T., Perote, J., Schmidt, U., Loos, M., 2009. Selfish-Biased Conditional Cooperation: On the Decline of Contributions in Repeated Public Goods Experiments. *Journal of Economic Psychology* 30 (1), 52–60. <https://doi.org/10.1016/j.joep.2008.04.005>.
- Neuteleers, S., Engelen, B., 2015. Talking Money: How Market-Based Valuation Can Undermine Environmental Protection. *Ecol. Econ.* 117 (September), 253–260. <https://doi.org/10.1016/j.ecolecon.2014.06.022>.
- Ottoni-Wilhelm, M., Vesterlund, L., Xie, H., 2017. Why Do People Give? Testing Pure and Impure Altruism. *Am. Econ. Rev.* 107 (11), 3617–3633. <https://doi.org/10.1257/aer.20141222>.
- Pagiola, S., Honey-Rosés, J., Freire-González, J., 2016. Evaluation of the Permanence of Land Use Change Induced by Payments for Environmental Services in Quindío, Colombia. *PLoS One* 11 (3), e0147829. <https://doi.org/10.1371/journal.pone.0147829>.
- Pagiola, S., Honey-Rosés, J., Freire-González, J., 2020. Assessing the Permanence of Land-Use Change Induced by Payments for Environmental Services: Evidence From Nicaragua. *Trop. Conserv. Sci.* 13 (January), 1940082920922676. <https://doi.org/10.1177/1940082920922676>.
- Pascual, U., Balvanera, P., Anderson, C.B., Chaplin-Kramer, R., Christie, M., González-Jiménez, D., Martín, A., et al., 2023. Diverse Values of Nature for Sustainability. *Nature* 620 (7975), 813–823. <https://doi.org/10.1038/s41586-023-06406-9>.
- de Quidt, J., Haushofer, J., Roth, C., 2018. Measuring and Bounding Experimenter Demand. *Am. Econ. Rev.* 108 (11), 3266–3302. <https://doi.org/10.1257/aer.20171330>.
- Rabe-Hesketh, S., Skrondal, A., Pickles, A., 2005. Maximum Likelihood Estimation of Limited and Discrete Dependent Variable Models with Nested Random Effects. *J. Econ.* 128 (2), 301–323. <https://doi.org/10.1016/j.jeconom.2004.08.017>.
- Rasch, S., Wünscher, T., Casasola, F., Ibrahim, M., Storm, H., 2021. Permanence of PES and the Role of Social Context in the Regional Integrated Silvo-Pastoral Ecosystem Management Project in Costa Rica. *Ecological Economics* 185 (July), 107027. <https://doi.org/10.1016/j.ecolecon.2021.107027>.
- Reichhuber, A., Camacho, E., Requate, T., 2009. A Framed Field Experiment on Collective Enforcement Mechanisms with Ethiopian Farmers. *Environ. Dev. Econ.* 14 (5), 641–663. <https://doi.org/10.1017/S1355770X09005178>.
- García-Amado, R., Luis, M.R., Pérez, and Sara Barrasa García., 2013. Motivation for Conservation: Assessing Integrated Conservation and Development Projects and Payments for Environmental Services in La Sepultura Biosphere Reserve, Chiapas, Mexico. *Ecol. Econ.* 89 (May), 92–100. <https://doi.org/10.1016/j.ecolecon.2013.02.002>.
- Rode, J., Gómez-Baggethun, E., Krause, T., 2015. Motivation Crowding by Economic Incentives in Conservation Policy: A Review of the Empirical Evidence. *Ecological Economics* 117 (September), 270–282. <https://doi.org/10.1016/j.ecolecon.2014.11.019>.
- Roe, B.E., Just, D.R., 2009. Internal and External Validity in Economics Research: Tradeoffs between Experiments, Field Experiments, Natural Experiments, and Field Data. *Am. J. Agric. Econ.* 91 (5), 1266–1271. <https://doi.org/10.1111/j.1467-8276.2009.01295.x>.
- Ryan, R.M., Deci, E.L., 2000. Self-Determination Theory and the Facilitation of Intrinsic Motivation, Social Development, and Well-Being. *Am. Psychol.* 67.
- Salk, C., Lopez, and G. Wong. 2017. 'Simple Incentives and Group Dependence for Successful Payments for Ecosystem Services Programs: Evidence from an Experimental Game in Rural Lao PDR'. [http://widgets.ebscohost.com/prod/customer-specific/s6689315/vpn\\_osnabrueck.php?url=http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edsosai&AN=edsosai.on1157292182&lang=de&site=eds-live&scope=site](http://widgets.ebscohost.com/prod/customer-specific/s6689315/vpn_osnabrueck.php?url=http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edsosai&AN=edsosai.on1157292182&lang=de&site=eds-live&scope=site).
- Salzman, J., Bennett, G., Carroll, N., Goldstein, A., Jenkins, M., 2018. The Global Status and Trends of Payments for Ecosystem Services. *Nature Sustainability* 1 (3), 136–144. <https://doi.org/10.1038/s41893-018-0033-0>.
- Schomers, Sarah, and Bettina Matzdorf. 2013. 'Payments for Ecosystem Services: A Review and Comparison of Developing and Industrialized Countries'. *Ecosystem Services, Payments for Ecosystem Services and Their Institutional Dimensions: Analyzing the Diversity of Existing PES Approaches in Developing and Industrialized Countries*, 6 (December):16–30. doi: 10.1016/j.ecoser.2013.01.002.
- Snilsveit, B., Stevenson, J., Langer, L., Tannous, N., Ravat, Z., Nduku, P., Polanin, J., Shemilt, I., Eyers, J., Ferraro, P.J., 2019. Incentives for Climate Mitigation in the Land Use Sector—the Effects of Payment for Environmental Services on Environmental and Socioeconomic Outcomes in Low- and Middle-income Countries: A Mixed-Methods Systematic Review. *Campbell Systematic Reviews* 15 (3), e1045.
- Stanley, T.D., Doucouliagos, H., 2014. Meta-Regression Approximations to Reduce Publication Selection Bias. *Res. Synth. Methods* 5 (1), 60–78. <https://doi.org/10.1002/jrsm.1095>.
- Stanley, T.D., Doucouliagos, H., Ioannidis, J.P.A., 2022. Beyond Random Effects: When Small-Study Findings Are More Heterogeneous. *Adv. Methods Pract. Psychol. Sci.* 5 (4), 25152459221120427. <https://doi.org/10.1177/25152459221120427>.
- Travers, H., Clements, T., Keane, A., Milner-Gulland, E.J., 2011. Incentives for Cooperation: The Effects of Institutional Controls on Common Pool Resource Extraction in Cambodia. *Ecol. Econ.* 71 (November), 151–161. <https://doi.org/10.1016/j.ecolecon.2011.08.020>.
- Vollan, B., 2008. Socio-Ecological Explanations for Crowding-out Effects from Economic Field Experiments in Southern Africa. *Ecol. Econ.* 67 (4), 560–573. <https://doi.org/10.1016/j.ecolecon.2008.01.015>.
- Vorlaufer, T., 2019. Effects of Double-Anonymity on pro- and Anti-Social Behavior: Experimental Evidence from a Lab in the Field. *J. Behav. Exp. Econ.* 81 (August), 216–225. <https://doi.org/10.1016/j.socec.2019.07.001>.
- Vorlaufer, T., Engel, S., de Laat, J., Vollan, B., 2023. Payments for Ecosystem Services Did Not Crowd out Pro-Environmental Behavior: Long-Term Experimental Evidence from Uganda. *Proc. Natl. Acad. Sci.* 120 (18), e2215465120. <https://doi.org/10.1073/pnas.2215465120>.
- World Bank. 2018. 'Evaluating the Permanence of Forest Conservation Following the End of Payments for Environmental Services in Uganda'. <http://documents1.worldbank.org/curated/en/867301589223292444/pdf/Evaluating-the-permanence-of-forest-conservation-following-the-end-of-payments-for-environmental-services-in-Uganda.pdf>.
- Wunder, S., 2015. Revisiting the Concept of Payments for Environmental Services. *Ecol. Econ.* 117 (C), 234–243.
- Wunder, S., Börner, J., Ezzine-de-Blas, D., Feder, S., Pagiola, S., 2020. Payments for Environmental Services: Past Performance and Pending Potentials. *Ann. Rev. Resour. Econ.* 12 (1), 209–234. <https://doi.org/10.1146/annurev-resource-100518-094206>.
- Zizzo, D.J., 2010. Experimenter Demand Effects in Economic Experiments. *Exp. Econ.* 13 (1), 75–98. <https://doi.org/10.1007/s10683-009-9230-z>.